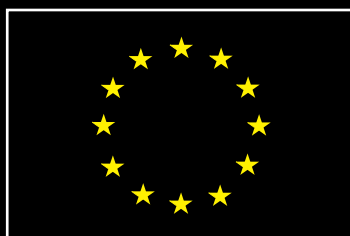


# EUROPEAN ECONOMY

EUROPEAN COMMISSION  
DIRECTORATE-GENERAL FOR ECONOMIC AND FINANCIAL AFFAIRS



REPORTS AND STUDIES

**LIBERALISATION OF NETWORK INDUSTRIES**  
Economic implications and main policy issues

---

**European Commission**

# **EUROPEAN ECONOMY**

**Directorate-General for Economic and Financial Affairs**

**1999**

**Number 4**



# **Liberalisation of network industries**

Economic implications  
and main policy issues

## **Abbreviations and symbols used**

### **Member States**

B	Belgium
DK	Denmark
D	Germany
WD	West Germany
EL	Greece
E	Spain
F	France
IRL	Ireland
I	Italy
L	Luxembourg
NL	The Netherlands
A	Austria
P	Portugal
FIN	Finland
S	Sweden
UK	United Kingdom
EU-9	European Community excluding Greece, Spain and Portugal
EU-10	European Community excluding Spain and Portugal
EU-12–	European Community, 12 Member States including West Germany
EU-12+	European Community, 12 Member States including Germany
EU-15+	European Community, 15 Member States including Germany
EUR-11	Group of 11 Member States participating in monetary union

### **Currencies**

ECU	European currency unit
EUR	Euro
ATS	Austrian schilling
BEF	Belgian franc
DEM	German mark (Deutschmark)
DKK	Danish krone
ESP	Spanish peseta
FIM	Finnish markka
FRF	French franc
GBP	Pound sterling
GRD	Greek drachma
IEP	Irish pound (punt)
ITL	Italian lira
LUF	Luxembourg franc
NLG	Dutch guilder
PTE	Portuguese escudo
SEK	Swedish krona
CAD	Canadian dollar
CHF	Swiss franc
JPY	Japanese yen
RUR	Russian rouble
USD	US dollar

### **Other abbreviations**

ACP	African, Caribbean and Pacific States having signed the Lomé Convention
ECSC	European Coal and Steel Community
EDF	European Development Fund
EIB	European Investment Bank
EMCF	European Monetary Cooperation Fund
EMS	European Monetary System
ERDF	European Regional Development Fund
Euratom	European Atomic Energy Community
Eurostat	Statistical Office of the European Communities
GDP (GNP)	Gross domestic (national) product
GFCF	Gross fixed capital formation
LDCs	Less developed countries
Mio	Million
Mrd	1 000 million
NCI	New Community Instrument
OCTs	Overseas countries and territories
OECD	Organisation for Economic Cooperation and Development
OPEC	Organisation of Petroleum Exporting Countries
PPS	Purchasing power standard
SMEs	Small and medium-sized enterprises
toe	Tonne of oil equivalent
:	Not available



# Contents

<b>Introduction and synopsis</b>	<b>5</b>
<b>PART A:    Liberalisation of network industries:              Economic implications and main policy issues</b> by Fabienne Ilzkovitz, Roderick Meiklejohn and Ulrik Mogensen under the direction of Jan Host Schmidt	<b>11</b>
Executive summary	15
1.     Introduction	19
2.     Concept and characteristics of network industries	21
3.     State and process of liberalisation	26
4.     Economic implications of liberalisation of network industries — telecommunications as an example	30
5.     Competition policy and regulation	36
6.     Services of general interest and the operation of the single market	42
7.     Lessons from the Community's perspective	47
8.     Conclusion	48
Bibliography	50
 <b>PART B:    Network industries and public service — study carried out for the European              Commission by the Institut d'Économie Industrielle, Université des Sciences              Sociales de Toulouse</b>	 <b>57</b>
 <b>Introduction</b>	 <b>67</b>



<b>Part One: Deregulation of network industries and the role of the public authorities</b>	<b>81</b>
I.    Network industries and network goods	81
II.   Government intervention in network industries	99
III.  Competition policy versus regulation	123
IV.   Pricing interconnection	132
 <b>Part Two: Universal service</b>	 <b>168</b>
I.    Universal service: definition, cost and financing	168
II.   Universal service in practice — some experiences	185
 <b>Part Three: Main policy issues in network industries</b>	 <b>198</b>
Telecommunications	198
Postal service	200
Electricity	202
Natural gas	204
Water	206
Urban transport	207
Air transport	209
Railway transport	211

<b>Tables</b>	1. Value added in the network industries as a percentage of GDP, 1994	22
	2. Average share of household budgets, 1994 (%)	22
	3. Some sectoral differences in network industries	23
	4. Country differences (electricity industry as an example)	24
	5. Ownership of network industries, spring 1998	25
	6. Degree of liberalisation in Member States	26
	7. Degree of competition in network industries, examples	26
	8. Liberalisation of telecommunications in the EU	31
	9. Main telecommunications directives and their implementation	51
	10. Network connection and supply times	52
	11. Fault repair: Targets and actual times	53
	12. Telecommunications penetration, 1996	54
	13. GDP per head and telecommunications, 1996	54
	14. Number of new jobs in 2005 resulting from liberalisation of telecom	55
	15. Main characteristics of network industries	96
	16. Bottlenecks, bypass and duplication possibilities	97
	17. Productive efficiency — comparative studies of public and private firms	119–120
	18. Efficiency of european railways	121
	19. Comparison of efficiency measures for railways	121
 <b>Graphs</b>	1. Driving forces for and resistance forces towards liberalisation and competition	27
	2. Prices for business calls and number of operators	32
	3. Residential telecom prices and number of operators	33
	4. Overall, allocative and technical measures of productive efficiency	115
	5. Technical measures of efficiency in input-output space	116



# Introduction and synopsis

## Why such a report on the liberalisation of network industries?

At the Cardiff European Council in June 1998, Europe's leaders underlined the important contribution of structural reforms to fostering growth and employment. This Council also asked for closer monitoring and coordination of structural reforms. To answer these requests, the Commission prepared two reports: the so-called Cardiff I and II reports analysing respectively the functioning of product and capital markets and economic and structural reforms in the EU. The basic objective of these reports was to launch a discussion on the structural reforms deemed appropriate to foster growth and employment and to ensure a smooth functioning of EMU. The implicit assumption is that regulation is an important element which affects the performance of markets and that excessive or inappropriate regulations can affect a country's economic performance and lead to a waste of resources. Therefore, for many countries, a way to increase growth is to improve the quality of regulation and economic policies.

The present publication has to be seen in the context of this 'Cardiff process' as it focuses on an important area of structural reforms taking place in the EU, i.e. the liberalisation of network industries. Its major objectives are to define the conceptual framework for analysing this liberalisation process, to illustrate its economic consequences and to discuss the resulting implications for the organisation and regulation of the liberalised industries.

Market liberalisation generally goes hand-in-hand with regulatory reforms, as the nature of regulation has to adapt over time to the new economic environment (globalisation, shifts in technology etc.). An improvement in the regulatory environment does not necessarily mean deregulation and it may also entail re-regulation provided that the new rules are better adapted to the

new economic environment. In a changing business environment, regulations that were appropriate in the past may no longer be so in the future. They may even prevent enterprises from competing successfully in the larger EU or world markets, causing a destruction of employment or productive capacity. Governments therefore have a responsibility to keep regulations under continuous review and assess their appropriateness in the current business environment.

This issue of regulation and its evolution over time is particularly important for the network industries where the respective roles of regulatory and competition authorities must evolve with the degree of liberalisation. These industries (telecommunications, postal services, air, rail and urban transport, electricity, gas and water) make up 5 % of EU GDP and employment but their economic importance is greater still because the price and quality of their outputs are important determinants of the growth and competitiveness of European industries, the operation of the single market and the European consumers' living standards. In these industries, market liberalisation and regulatory reforms are expected to enhance static (better resource allocation) and dynamic (innovation, new products and technologies) efficiency of firms. This should lead not only to cost and price reductions but also to quality improvements and increased demand. However, changes brought about by competition also imply adjustment costs because deregulation can be accompanied by employment losses in the short term in the sectors concerned.

Therefore, the main challenges are today to improve the public acceptance of this liberalisation process, on the one hand, and to combine efficiency, competition and the provision of services of general interest on the other. Acceptance of the liberalisation process will be greater if its positive effects are clearly identified and understood by industrial users and consumers. It is also important that the benefits of liberalisation should be

shared fairly among the users in the form of lower prices, wider choice and better quality. Effective and independent national regulatory authorities are therefore essential. However, in many Member States the regulatory authorities have been set up only recently and have yet to develop adequate expertise. There is also an important role of central coordination of national regulators to be played at the European level.

The second challenge concerns obligations to provide services of general interest. Many services provided by the network industries have come to be regarded as essential public services which should be widely available at 'affordable' prices and with assured quality. As the free operation of the market may not always meet these requirements, Member States may impose public service obligations. However, the definition of these obligations, their financing and the selection of the enterprises charged with carrying them out should not create unnecessary distortions of the market or impede the process of integration. In particular, it is very important that the services of general interest are clearly defined and that the conditions of their provision are transparent so as to avoid compromising the benefits flowing from an internal market.

These two challenges and other policy implications of the liberalisation of the network industries are discussed in detail in the present publication, which consists of two parts. The first part, written by staff of the Directorate-General for Economic and Financial Affairs, presents an overview of the present situation, the economic implications of the liberalisation of the network industries and the main policy issues. The second part consists of a wide-ranging study carried out for the Commission by the Institut d'Économie Industrielle of the Université des Sciences Sociales, Toulouse (IEI).

## **Part A: Economic implications and main policy issues**

The first part starts by discussing the concept and characteristics of the network industries. The network industries are an indispensable element of the economy, providing basic services to all other sectors. The typical feature of the network industries is that the network infrastructure usually constitutes a heavy fixed cost, implying substantial economies of scale and some element of natural monopoly. However, the importance of the fixed infrastructure costs varies greatly between the

industries, being lowest in the local public transport and postal sectors and highest in the electricity and gas industries. Section 2.3 compares the network industries in terms of a number of characteristics, such as their growth prospects, capital intensity, the extent of the natural monopoly and the importance of public service obligations. The next section compares the organisation of the industries in different Member States, showing that wide variations exist even within the same sector. For example, the electricity industry in France is publicly owned, centralised and fully vertically integrated, while in the UK the industry has been privatised, creating a large number of companies operating at different stages in the supply chain.

Chapter 3 considers the process of liberalisation and, in particular, the forces which are driving this process and those which are resisting it. The Community's efforts to complete the single market are amongst the most important driving forces, but technological change, pressure from dissatisfied customers and potential competitors are also major factors. In addition, tighter budgetary discipline has led Member States to seek ways of reducing expenditure on subsidies and raise revenue from the sale of public enterprises. On the other hand, liberalisation poses a number of risks. Some consumers fear that reform may lead to higher prices or lower quality, while employees fear losing their jobs or being obliged to accept inferior conditions of employment. At the same time, managers of some incumbent companies resist change either because they do not want to lose their monopoly position or because they fear that the entry of competitors will leave them with 'stranded costs'. The impact of liberalisation on employment is of special importance, in view of the persistently high level of unemployment in the Community. Section 3.4 discusses this problem. Although the overall, long-term impact on employment may well be positive, it is undeniable that short-term frictional unemployment may be created and appropriate measures need to be taken to alleviate this problem. However, liberalisation is only one of the factors affecting employment in these sectors. Technological change, in particular, has a substantial impact, by raising productivity, making old skills redundant and causing shifts in demand, e.g. from postal services to telecommunications.

Chapter 4 looks in more detail at liberalisation in telecommunications, the sector where the process is farthest advanced. It compares progress in the different Member States and examines the impact on prices, the

quality of service and employment. Unfortunately, the available information does not allow us to draw any definite conclusions about other effects, for two main reasons. First, even in this sector, liberalisation is far from complete in most Member States. Second, it is very difficult to distinguish between the effects of liberalisation and those of the rapid technological progress which has occurred in this sector in the last decade. Nevertheless, it seems clear that there is a link between liberalisation and lower prices. Regarding employment, there are short-term employment losses as the employment created by the new operators is insufficient to offset the lay-off by incumbents. However, the long-term prospects are better as the reduction in prices will increase demand and spread into other sectors, leading to competitiveness gains for the whole economy.

Chapter 5 discusses the problems of introducing and policing competition in the network industries. The first three sections of the chapter discuss the roles which can be played by sector-specific regulatory bodies, on one hand, and economy-wide competition authorities on the other. The main conclusion is that specific regulation will continue to be necessary for the foreseeable future but that it should operate within a general framework provided by competition law. Furthermore, the nature and scope of regulatory intervention should be adjusted progressively as each industry becomes more open to competition. Section 5.4 discusses the question of the appropriate geographical level of regulation, taking the principle of subsidiarity into account. It is argued that, as a general rule, regulation should be implemented by national, regional or local bodies. However, where there are important cross-border spillovers, as in some areas of telecommunications or in air traffic control, a Community or international authority may be necessary. The Community can also play a role in helping national authorities to coordinate their activities, promoting the exchange of information and the diffusion of best practice. Within a Community framework designed to ensure that single market principles are respected, the definition of regulatory principles is best done at the national or sub-national level.

Chapter 6 examines the question of 'services of general interest'. All Member States impose obligations on their network industries to provide specific services which are deemed to be of general interest or to take account of strategic objectives such as security of supply. In addition, the legislation often stipulates that certain services must be provided at 'affordable' prices or that

uniform prices must be charged throughout the country. The underlying policy objectives are usually those of regional policy or social welfare. When the imposition of such obligations imposes large costs on the undertakings concerned, important distortions can arise. This chapter concludes that an appropriate solution to this problem would be to allocate responsibilities for providing services of general interest by means of an open public tendering procedure and to compensate for the costs through direct subsidies. The last section of the chapter points out that the definition of these obligations and the ways in which they are allocated and financed have implications for the single market. It is essential to avoid national measures which may create barriers to the entry of firms from other Member States or shift the costs of the services of general interest to other countries.

Chapter 7 draws some general lessons from the Community's point of view. It emphasises the fundamental role of Community competition law and policy and identifies important roles for the EU in coordinating national regulatory bodies and promoting liberalisation and the integration of European networks.

The final chapter of the first part summarises the main conclusions. Important efficiency gains can be expected from the liberalisation of the network industries, but the adjustment costs may in some cases be significant, particularly in terms of short-term job losses. Technological developments and changes in ownership and market structure pose major challenges to both regulators and managers. These challenges demand considerable flexibility both in the industries themselves and in the regulatory systems.

## **Part B: The conceptual and theoretical framework**

The IEI study provides a detailed conceptual and theoretical framework, which served as the main basis for the preparation of the first part of this publication. The study opens with an introduction which summarises the economic characteristics of the network industries, discusses the rationale and mechanisms of public intervention and provides an overview of the concept of universal service and the problems which it poses. The last section of the introduction briefly discusses the main policy issues in each of the industries.

Part One of the study (Deregulation of network industries and the role of the public authorities) focuses on four issues: basic definitions and concepts, government intervention, competition versus regulation and inter-connection pricing.

Chapter I presents the basic features of networks (nodes and interconnections) and their economic features: the relative importance of (fixed) infrastructure and (variable) operating costs, economies of scale, scope and density, externalities, etc. These basic economic concepts are used to characterise the typical networks of different industries (energy, transport and communications). On the basis of these concepts, the study identifies the borderlines between natural monopolies, mixed activities and competitive areas in these network industries.

Chapter II goes on to discuss the modes of government intervention (legally and/or economically) including contractual and corporate governance issues. It discusses the rationale for government intervention on the basis of a range of arguments including market imperfections, externalities, merit goods, income and regional redistribution objectives, etc. In an appendix a survey of studies testing the relative performance of public and private firms shows that performance cannot be systematically related to ownership. Indeed, it is clear that privatisation is neither a sufficient nor always a necessary condition for improving the efficiency of the network industries.

Chapter III presents the distinction between the traditional sectoral regulatory approach to network industries and the wider competition policy approach as applied by economy-wide competition authorities. The study argues that the differences in these approaches are substantial and that during the current wave of structural reform and liberalisation both types of government intervention are likely to be necessary and complementary.

Chapter IV introduces the reader to the complex issue of interconnection. Unless ownership and management of the network infrastructure are separated from the provision of final services, the incumbent operator (the former monopolist) typically controls essential facilities (bottleneck infrastructure), to which its competitors need access. The conditions under which other operators are allowed access to these facilities, in particular the pricing of access, are major factors determining the

degree of competition. Building on past experience, for example that of OFTEL (the British telecom regulator), the chapter discusses different pricing approaches including fully distributed costs (FDR), the efficient component pricing rule (ECPR) and the 'OFTEL rule'. The chapter goes on to discuss the efficient conditions for entry, bypass and competition in different environments, for instance under one-way and two-way access.

Traditionally, regulators have tried to control a number of different prices individually, including not only various types of access charges but also prices for final services. This study advocates the application of a global price cap, which limits the aggregate price of a 'basket' of services provided by the operator. An incumbent operator subject to a global price cap will set the price of each type of service at the most economically efficient level, provided that the incumbent has no incentive to exclude competitors by implementing a 'price squeeze' (combining high access charges with low final prices). It may be necessary to supplement the global price cap with other controls when the latter condition is not met.

Part Two (Universal service) is devoted to the concept of universal service obligation. Chapter I discusses definitions, costs and financing of universal service obligations (USOs) and the rationale for this type of government intervention. The authors distinguish between public service and universal service. They use the former term to refer to a wide range of public policy objectives, including security of supply and protection of the environment, while the term universal service is used in this study to refer only to a limited set of objectives: the provision of services of a specified quality, available to all at 'affordable' prices. Building on the discussion in Part 1, Chapter II, the authors discuss network externalities, redistribution and merit goods as possible economic arguments for USOs. They conclude that arguments based on network externalities provide only a weak justification for these obligations. The third section of Chapter I discusses the problems of measuring the cost imposed on the operator by USOs and analysing the social costs and benefits of the measures. The last section examines methods of financing the costs both in a monopolistic sector and in a liberalised industry. Chapter II explains how USOs are defined and financed in some European countries, the USA and Japan. This chapter also outlines the relevant Community rules in some of the sectors.

Finally, Part Three (Main policy issues in network industries) summarises for each of the network industries covered by the study the main policy challenges for the coming years. Though the concepts used and the questions currently posed in the different network industries are somewhat similar, the presentation clearly shows that the technical, physical, organisational and economic conditions are often sector- and/or country-specific. This implies that a synthesised approach does not succeed in capturing satisfactorily the many challenges and issues facing network industries in the com-

ing years. In most of the sectors, however, access to the bottleneck infrastructure is a major issue. The authors argue that even in the telecommunications sector this issue has not yet been satisfactorily resolved. Amongst the sector-specific problems, the authors stress the employment impact of liberalisation of the postal service. The many other questions raised include the problem of 'stranded costs' in electricity generation, the need to ensure that there is some coordination of investment in the gas and electricity industries and the environmental implications of liberalisation in the water industry.





# Part A

## Liberalisation of network industries: Economic implications and main policy issues <sup>(1)</sup>

Written under the direction of Jan Host Schmidt by Fabienne Ilzkovitz, Roderick Meiklejohn and Ulrik Mogensen, Directorate E, Directorate-General for Economic and Financial Affairs.

Thanks are due to a number of colleagues in the Directorate for invaluable assistance and advice.

---

<sup>(1)</sup> This publication does not necessarily reflect the views of the European Commission.



# Contents

<b>Executive summary</b>	<b>15</b>
<b>1. Introduction</b>	<b>19</b>
<b>2. Concept and characteristics of network industries</b>	<b>21</b>
2.1. Concept of network industries	21
2.2. Economic importance	21
2.3. Sectoral characteristics	23
2.4. Country characteristics	24
<b>3. State and process of liberalisation</b>	<b>26</b>
3.1. Current state of liberalisation by country and sector	26
3.2. Driving forces for liberalisation and competition	27
3.3. Resistance towards liberalisation and competition	28
3.4. Liberalisation and employment	28
<b>4. Economic implications of liberalisation of network industries — telecommunications as an example</b>	<b>30</b>
4.1. Degree of liberalisation in the Member States	30
4.2. Liberalisation and prices	32
4.3. Liberalisation and quality	33
4.4. Liberalisation and penetration	34
4.5. Impact on employment	34
<b>5. Competition policy and regulation</b>	<b>36</b>
5.1. Differences between competition policy and regulation	36
5.2. The need for regulation	37
5.3. The relationship between regulation and competition policy	38
5.4. Subsidiarity and the level of regulation	39
5.5. Access to the bottleneck infrastructure	39
5.6. Competition in the market and competition for the market	40
<b>6. Services of general interest and the operation of the single market</b>	<b>42</b>
6.1. Definition of OSGI	42
6.2. <i>Raison d'être</i> of OSGI	42
6.3. The provision and financing of services of general interest	43
6.4. OSGI and competition	44
6.4.1. The Treaty provision	44
6.4.2. Is competition consistent with OSGI?	45
6.5. OSGI and the single market	45

<b>7. Lessons from the Community's perspective</b>	<b>47</b>
<b>8. Conclusion</b>	<b>48</b>
<b>Bibliography</b>	<b>50</b>
<b>Annex</b>	<b>51</b>

# Executive summary

1. Network industries are characterised by the delivery of products or services to final customers via 'a network infrastructure' linking upstream supply with downstream customers. The eight network industries analysed here are telecommunications, postal services, energy (electricity and natural gas), transport (urban, air and railways) and water. The network structure is typically rather costly to establish and embodies substantial fixed costs, implying lower average production costs for larger scale production. This is the main reason explaining the existence of natural monopolies in these industries. However, technological progress has reduced the extent of natural monopolies and has changed the nature of regulation in network industries. For example, as competition in these industries is introduced in an environment in which there exist large and often dominant incumbents, regulation is necessary in the early stages of liberalisation to protect new entrants (see point 8).
2. Network industries account for more than 6 % of the EU's GDP. However, their economic importance is larger than that indicated by this ratio as the price and quality of their outputs is essential for growth and competitiveness of European industries, for the well-functioning of the internal market and for the standard of living of European consumers. The economic importance of network industries differs across sectors. For example, communications represent approximately 2 % of GDP while the sector of urban transport is much smaller (less than 0.4 % of GDP). For a given sector, the GDP share can also vary significantly between countries: the GDP share of communications services ranges between 1.8 % in Italy and 3.3 % in Luxembourg.
3. Common characteristics can be found among the network industries: existence of natural monopolies, dominance of incumbents, obligations to provide services of general interest, importance of regulation, etc. However, despite these common characteristics, significant differences exist between sectors. The network industries cover very diversified industries in terms of growth (telecommunication versus railways), capital intensity (air transport versus postal services), degree of internationalisation (postal services versus water supply) and degree of competition (monopoly in railways, except in the United Kingdom, versus effective competition in non-reserved postal services). These industries are also organised very differently across countries. In some countries, network industries have traditionally been organised in centralised, vertically integrated public monopolies (France) while in others, private regional or local companies play a significant role (Germany). These differences explain that it is difficult to analyse these industries within an overall generic framework. However, some policy issues overlap in many cases.
4. One important policy issue is the liberalisation process in the network industries. Again, the situation varies greatly across countries and sectors. Broadly speaking, one can consider that the liberalisation process is most advanced in the United Kingdom and the Nordic countries and least advanced in southern Europe. The degree of competition is higher in the telecommunications industry and in the non-reserved postal services and lower in the water and railways industries. Between these extremes, one finds the energy sectors and other transport services.
5. The liberalisation process in network industries is very gradual in Europe because there are simultaneously driving forces in favour of this process and elements of resistance against it. The main driving forces are progress in European integration, technological developments, which have per-

- mitted a decline in fixed costs, reducing the *raison d'être* of monopolies, and the deterioration of public finances which has reduced resources to finance the huge investments necessary in many network industries. In addition, the importance of competition as a factor bringing down prices and increasing innovation has been more widely recognised, while the natural monopoly and competitive elements are more clearly distinguished. The main forces of resistance are the lobbying of incumbents that fear the loss of their privileged position, the resistance of employees and trade unions because of the risk of job destruction and equity considerations making it necessary to ensure access to essential services at affordable prices.
6. It is very difficult to measure the economic impact of liberalisation for two main reasons. First, it is difficult to isolate this impact from other factors having affected the performance of network industries, such as technological progress. Second, the period of time that has elapsed since their liberalisation is too short. However, the analysis of the telecommunication industry has led to some interesting observations. It has shown that liberalisation goes hand-in-hand with lower prices and changes in the price structures. Even the anticipation of liberalisation may lower prices. Due to lack of data, no conclusion can be drawn about the link between quality of services and the progress of the liberalisation. Regarding employment, there may be short-term employment losses in the sectors concerned if the employment created by the new operators is insufficient to offset the lay-off by incumbents. However, the long-term prospects are better, as the reduction in prices will increase demand and spread into other sectors, leading to competitiveness gains for the whole economy.
  7. Liberalisation and the introduction of competition in network industries have increased the role of competition authorities in these sectors. This does not mean that the regulation of network industries is unnecessary but rather that there is a need for a great coordination between regulatory and competition authorities. The differences between competition authorities (CA) and regulatory agencies (RA) are not clear-cut. They are rather a question of degree along the following scales: sectoral scope of control (more sector-specific for RA), objectives (broader for RA), timing of oversight (*ex post* and lengthy for CA, *ex ante* and more rapid for RA), relationship with the industry (long-term relations and bigger influence of the industry for RA) and information requirements (specific expertise for RA). The division of responsibilities is therefore that competition and single market law provides the general framework within which regulation operates while the regulators, with their specific knowledge and more rapid and flexible procedures, are responsible for applying the sector-specific rules.
  8. The respective roles of regulatory and competition authorities must also evolve over time with the degree of liberalisation. A study carried out by the CEPR and the SNS (Bergman et al., 1998) suggests a stylised evolution of the network industries in three phases. In a first step, when the liberalisation process has not yet started, the regulation is mainly concerned with the prevention of monopoly abuse in retail markets by dominant incumbent firms. In a second step where the monopolies are broken up, regulatory activity increases. Regulatory authorities have to promote entry and prevent foreclosure. This implies that they will have to resolve problems of interconnection and access to the existing network infrastructures for new entrants. In this step, mechanisms are also put into place to preserve services of general interest and protect consumers. In a third step, when competition is more effective, the need for sector-specific rules diminishes but conditions and pricing of access to networks will continue to involve both regulatory and competition authorities, which may also have a continuing role to play in guaranteeing the provision of services of general interest.
  9. Many methods have been proposed to control access prices and a consensus seems to emerge in favour of a cost-based pricing rule, although the information requirements are heavy. Another problem is that, in the network industries, there are a number of markets which cannot be opened to competition because of bottleneck infrastructure or obligations to provide services of general interest. For example, in the sector of air transport, the possession of a 'slot' at a peak time can give an operator a high market power when only a

small number of slots is available. One way of dealing with this problem of inadequate competition in the market is to create competition for the market by auctioning franchises. For example, this approach has been adopted at the European Union level for allocating public service obligations to airlines and in Denmark for regional and local bus services.

10. The design and the geographical level of the regulatory framework will differ according to the industry and the nature of the problems addressed by specific rules. For example, the existence of important cross-border externalities, as in telecommunications or air traffic control, would justify the existence of rules at the Community or international levels. This institutional design must also be dynamic in nature, depending on the degree of liberalisation reached by a particular network industry. It is therefore important to ensure that the performance of regulatory agencies is periodically assessed and can evolve over time. However, although sector-specific regulation might be necessary, the responsible authority does not necessarily have to be sector-specific. In the EU context, it also seems necessary that the Commission play a role of central coordination to avoid that national regulations hinder economic integration.
11. Obligations to provide services of general interest (OSGI) occupy a prominent place in the debate regarding the liberalisation of network industries in the EU. The objective of these obligations is to make sure that everyone has access at an affordable price to a defined set of services which are considered essential. Various arguments can be given to justify OSGI: network externalities, i.e. when the benefits from using a network depend on the number of users, provision of public goods, or regional policy, because the provision of basic public services in small villages may prevent the decline of rural areas and redistribution concerns. However, the OSGI may also be the outcome of a political process or result from regulatory capture by some pressure groups and it is difficult to judge whether an OSGI is mainly motivated by social welfare considerations.
12. The importance which the European Union attaches to services of general interest is highlighted in the new Article 16 (former 7d) of the EC Treaty. The traditional way of financing these services is via cross-subsidisation from profitable services to loss-making services. It is often argued that a reserved area (legal monopoly) is necessary to protect the incumbent utility supplying these services against competitors only interested in offering the profitable services. Therefore, OSGI are often used as an argument against competition. However, the introduction of competition can also speed up the introduction of technology and thereby lead to price decreases, thus contributing to 'general interest' objectives. Although competition can help to achieve these objectives, the existence of high cost or low-income consumers implies that some regulation is still needed.
13. Regarding the financing of OSGI, it is not possible to define a single mechanism that would be appropriate in all sectors and all countries. Therefore, the costs of OSGI may be financed in a variety of ways. However, financing through a universal service fund to which all firms of the industry contribute has been considered in some countries and offers some advantages: it is transparent and does not provide incentives for inefficient by-pass. Problems with this approach are connected with the estimation of the incumbent's costs of providing the services in question, the allocation of these costs across the firms in the industry and the fact that it reinforces the position of the incumbent by making it a special player in the industry. In addition, universal service fund contributions might work as an entry barrier. A way to estimate and minimise costs could be to introduce a system of auction for the provision of services of general interest. In economic terms, the least distorting way of financing OSGI would be through direct government subsidies. An appropriate solution could then be to combine the system of auction with direct subsidies.
14. The concept of services of general interest is, to a large extent, country and sector specific and principally a matter for the Member States. But different definitions in different Member States could create competition problems and hinder the operation of the single market. In particular, in sectors where networks develop across national boundaries, such as telecommunications and energy, it is



- necessary to avoid competition between regulators and discrimination between national and foreign firms. In principle, Article 90 of the Treaty states that competition cannot be distorted on the pretext of fulfilling these obligations but general economic interest may override competition principles in cases of conflicts. However, the precise circumstances in which it is justified that undertakings' operating services of general economic interests are exempted from competition rules are not defined.
15. There is an important role of central coordination and of delivery of information to be played at the level of the European Union. For example, there is a need to provide guidance to Member States in relation to their OSGI. More precisely, the European Union should facilitate liberalisation and promote European integration of network industries. It should facilitate fair and efficient competition through competition policies. Finally, it should also coordinate national regulatory bodies and contribute to a periodic assessment of their tasks, in particular their power of enforcement.
  16. In conclusion, this paper has shown that network industries, which have been traditionally sheltered from competition and operated within national or local boundaries, have experienced important changes over the recent period. These changes should lead to efficiency gains in the medium term but they also imply adjustment costs in the short term. These adjustment costs include short-term employment losses, even if part of these losses is due to technological developments. Shifts in technology, ownership and market structure also create big regulatory and organisational challenges, as the nature of regulation and the kind of organisation have to evolve over time with the degree of liberalisation and need to take into account objectives of social welfare such those pursued by OSGI. From the EU's point of view, the main challenges are to provide guidelines, based on sound principles, which will facilitate liberalisation, increase its social acceptability and bring about greater consistency in regulatory practice throughout the Community.

# 1. Introduction

In the light of progressing liberalisation and structural reforms in network industries, the present report analyses various economic issues related to these industries. Eight network industries are considered here: telecommunications, postal services, energy (electricity and natural gas), transport (urban, air and railways) and water. This report takes a horizontal approach across these eight industries, while respecting the differences in sector-specific characteristics. The main advantage of this horizontal approach is that it permits the identification of major policy issues which are common to these eight network industries.

This report has benefited from the research carried out in this area by several research institutes. In particular, three reports have been particularly useful in its preparation: the report of the CEPS Working Party on utilities (see CEPS, 1996), the report on network industries and public service prepared by the Institut d'Economie Industrielle of the University of Toulouse (see Part B) and the CEPR/SNS report on the liberalisation in network industries (see Bergman et al., 1998).

The major objectives of this report are to illustrate the economic consequences of the liberalisation process in network industries observed in Europe and to analyse the resulting implications for the organisation and regulation of the liberalised industries. Liberalisation of network industries implies that competition becomes more effective in these industries as it is now recognised that it is a more powerful means of achieving efficiency than monopoly. But liberalisation is not synonymous with deregulation. Rather, the nature of regulation evolves over time with the degree of liberalisation. Organisation and regulation of the network industries also reflect the extent to which non-economic objectives are pursued in the form of, for instance, obligations to provide services of general interest.

Similarly, privatisation is different from liberalisation and is related to the change of ownership from public

authorities to private operators <sup>(1)</sup>. Liberalisation is generally accompanied by privatisation and there has been a gradual shift towards privatisation in network industries since the late 1980s. However, even today, there remains considerable public ownership in these industries. Economic theory is generally agnostic as to whether private ownership is superior in efficiency terms to public ownership. But empirical studies have tended to favour private ownership in some of the network industries. However, privatisation is not a sufficient condition to improve efficiency if it is not accompanied by liberalisation. This report focuses on liberalisation and does not analyse the consequences of privatisation.

This report starts by defining the concept of network industries and by describing their characteristics and economic importance. It highlights that the major justification for monopoly in these industries — the high fixed costs for the network structure — has in many cases lost its relevance with technological developments. This explains why competition has been introduced to a greater or lesser extent in these industries and has gathered momentum with the recent liberalisation process. A second section of this paper attempts to give a broad view on the state of liberalisation of these industries by country and by sector. It also identifies the driving forces in favour of this liberalisation and the factors of resistance against it. The telecommunication industry is used as an example to illustrate the possible impact of liberalisation in terms of prices, quality and employment.

In the medium term, increased competition should improve productivity and lead to lower prices in the

---

<sup>(1)</sup> It follows from Article 295 of the EC Treaty that the European Commission should be neutral vis-à-vis private or public ownership of network industries.

network industries, thereby stimulating competitiveness of European enterprises. However, much regulatory oversight will be required in these industries. The respective roles of competition and regulation authorities and their evolution over time with the degree of liberalisation are discussed in a fourth section of this report.

The price and quality of outputs supplied by network industries also affect the standard of living of European consumers. This is a reason why equity considerations

are taken into account when liberalising these industries. Therefore public service obligations are imposed by public authorities on undertakings rendering network services. The concept of obligations to provide services of general interest developed to describe these obligations is presented in the fifth section of this report. It also shows that some coordination is necessary between European regulators to avoid that these obligations hinder the operation of the single market. Finally, this paper concludes by drawing some policy recommendations from the Community's perspective.

## 2. Concept and characteristics of network industries

### 2.1. Concept of network industries

Network industries are characterised by the delivery of products or services to final customers via ‘a network infrastructure’ linking upstream supply with downstream customers <sup>(1)</sup>. The network is a structure of lines and nodes with a limited capacity that determines the direction (one or two way) and distribution of the network services. The network structure is typically rather costly to establish and embodies substantial fixed costs that determine economies of scale within the capacity of the network. Hence, doubling of and separation of networks is normally not economical. This would reduce the flexibility of the system and hinder exploitation of economies of scale. Natural monopolies are therefore an essential element of the network industries.

The natural monopoly element (the essential facility) however, might be more or less well defined and by-pass might be more or less possible. In the postal services, the incumbents’ logistic infrastructure can rather easily be by-passed whereas electricity provision is normally impossible without using the electricity grid.

These natural monopolies, however, depend on technical foundations <sup>(2)</sup>. As technology changes, the natural monopoly changes — both as a consequence of changes in the technology in the specific sector but also because developments of technologies in other areas. New forms of efficient small-scale production have emerged (electricity), the costs of connecting calls have decreased significantly (telecom) and niche production has been developed. Hence, due to technological

progress, the extent of the natural monopolies has been revised in recent years and new entrants have been allowed to provide services in areas, which were previously reserved for incumbents. In addition, the way in which natural monopolies could best be regulated in order to ensure efficient development of the networks has to be revised as well.

Network industries often deliver services or products which are essential for the general public and for the business sector (like energy, communication and transport). These services are consumed by everyone and are in addition very important for the competitiveness and development of other industries such as manufacturing, business services, research and development, etc. Thus, the economic importance of these sectors goes far beyond their share of GDP <sup>(3)</sup> and employment.

In addition, some network industries produce goods and services corresponding to basic needs to which everyone should have access. Consequently, obligations to provide services of general interest have been developed requiring the network industries to provide services even under circumstances where it is not economically profitable (see also Section 6).

### 2.2. Economic importance

Network industries make up more than 6 % of European GDP and employment. All sectors of the economy, including private companies, consumers and the public sector, use the services of the network industries. Therefore, they are important for European growth and the competitiveness of industry, since transport, energy and communications are essential for the well functioning of the internal market. The exact

<sup>(1)</sup> When a firm operates both the initial supply and the intermediate distribution or service provision it is vertically integrated. This definition refers to the CEPR report (see Bergman, 1998).

<sup>(2)</sup> Monopolies can however also depend on demand conditions. For instance, in a small village there is only room for one swimming pool, whereas in a large city several swimming pools can be economical even though the technology is the same.

<sup>(3)</sup> GDP ratios range from approximately 0.4 % for urban transport to 2 % in telecom in most countries.

impact, however, is difficult to quantify <sup>(1)</sup>. Little doubt remains that the potential of the internal market in Europe is handicapped by the present patchwork of transport, energy and telecom infrastructures.

<sup>(1)</sup> There is a general debate concerning the impact of public infrastructure investments on economic growth and econometric studies have been carried out on the subject. The results, however, are very sensitive to the particular cases and methods used and range from no impact at all to explaining one sixth of general economic growth.

Statistics covering these industries are poor and should be interpreted with caution. Table 1 gives figures on value added in the network industries as a percentage of GDP.

Eurostat conducts surveys on the composition of household expenditures in all Member States. Table 2 presents the share in household budgets of some services provided by network industries. The shares of all presented sectors (not all network industries) varies

*Table 1*

**Value added in the network industries as a percentage of GDP, 1994**

	Communications (inc. post)	Electricity and gas	Water	Urban transport	Air transport	Railways
Belgium	1.8	4.1		n.a.	n.a.	n.a.
Denmark	1.9	1.6	0.1	n.a.	n.a.	n.a.
Germany (West)	2.3	2.0	0.2	n.a.	0.7	0.40
Greece	n.a.	2.1		n.a.	n.a.	n.a.
Spain	2.1	2.3	0.3	0.2	0.9	0.01
France	2.3	2.3		0.3	0.9	0.50
Ireland	2.3	n.a.	n.a.	n.a.	n.a.	n.a.
Italy	1.8	5.9		n.a.	0.5	n.a.
Luxembourg	3.3	1.6		n.a.	n.a.	n.a.
Netherlands	2.1	1.4	0.3	0.04	1.5	n.a.
Austria	2.5	2.5	0.2	n.a.	n.a.	n.a.
Portugal	2.3	3.9		n.a.	n.a.	n.a.
Finland	2.0	2.1	0.3	n.a.	n.a.	n.a.
Sweden	2.1	2.3	0.4	0.1	n.a.	0.22
UK	2.7	2.2		n.a.	1.5	n.a.

Sources: IEI Study, OECD national accounts, national accounts of Germany and the UK.

*Table 2*

**Average share of household budgets, 1994**

	Electricity	Gas	Railway transport	Air transport	Postal services	Telecom services	Total
Belgium	1.9	1.3	0.3	0.2	0.1	1.6	5.4
Denmark	2.3	0.7	1.3	0.3	0.1	1.6	6.3
Spain	1.9	0.8	0.4	n.a.	0.0	1.4	4.6
Italy	1.5	2.1	0.3	n.a.	0.1	1.8	5.7
Netherlands	1.4	2.4	0.7	0.2	0.2	1.5	6.2
Portugal	2.7	0.7	0.3	0.1	0.0	2.0	5.8
Sweden	3.0	0.3	0.2	0.1	0.1	n.a.	3.8 <sup>(1)</sup>
UK	2.4	1.9	0.5	0.4	0.2	1.7	7.1

<sup>(1)</sup> Excluding telecommunications.

Source: Eurostat household budget survey. Unfortunately data are not available for Germany and France. No information on urban transport and water is available.

between 3.8 % in Sweden to 7.1 % in the UK. According to these consumer surveys, network industries receive an increasing share of consumers' budgets. Unfortunately, no data are available for Germany and France. The share of each item in the household budget of course depends on a whole range of factors including the price of the services, their availability, tariff structures, the average size of households in each country, the income level, geography, climate, etc.

However, it is interesting to observe the following. In most sectors, shares are fairly even throughout Europe, especially for telecom services. In the gas sector there are, however, large discrepancies between Member States reflecting the variations in availability. Sweden has a large share of expenditures on electricity in spite of cheap water-generated electricity. This is partly due to a heavy reliance on electric heating, colder weather and high degree of second home ownership in Sweden. The Netherlands have high levels of gas expenditure and low levels of electricity expenditure due to the highly developed gas industry in the country. Denmark, apparently, has either a high consumption of rail transport or very expensive train tickets. In the UK, air transport expenditures are a relatively large share of household budgets.

### 2.3. Sectoral characteristics

Some common characteristics can be found among network industries: existence of natural monopolies, domi-

nance of incumbents, obligations to provide services of general interest, regulation, etc. In spite of these common characteristics, important discrepancies between sectors remain. The network industries cover very diversified industries from high-growth, highly profitable sectors to stagnating sectors dependent on government subsidies.

The telecom sector, for instance, has seen extremely high growth in the last decade and the continuing development in the sector is likely to influence all parts of our societies and lead us towards the information society. The railway sector, on the contrary, has been stagnating for years and depends heavily on State subsidies. Differences in demand conditions are therefore very important not just for the sectoral characteristics but also for the potential of the liberalisation process.

Network industries have traditionally been quite capital intensive compared to other industries. Hence, the traditional public legal monopolies served to protect the relatively large investments embodied in infrastructure, etc. Nonetheless, within the group of network industries substantial differences in capital intensity can be observed. Energy sectors, for instance are very capital intensive whereas postal services are less.

Some sectors are internationally integrated like postal services, gas and electricity, where services are fairly standardised and networks are integrated across bor-

Table 3

#### Some sectoral differences in network industries

	High extreme	Low extreme
Economic importance/growth prospects	Telecom high growth sector 8 % p.a., air transport	Railways
Capital intensity	Air transport, telecom, electricity, gas	Postal services, urban transport
Extent of natural monopoly	In telecom the natural monopoly element is shrinking due to the emergence of alternative infrastructures.	In the water sector there is no alternative to existing network, which is a natural monopoly.
Technical compatibility	Postal services — standard format throughout Europe, electricity	Railways — different rail gauge, tracks, difference in power supply, etc.
Network structure	Two-way: Telecom, postal services, transport.	One-way: electricity, gas, water.
Capacity to store production	Gas and water: storable.	Electricity: non-storable.
Frequency of supply	Instant: telecom, gas, and electricity.	Variable: transport, postal services.

Note: By one and two-way network structure we mean that the consumers are able to use the network for both receiving and sending services. In the electricity grid it is possible to send power in both directions. However, the electricity grid cannot be characterised as a two-way network as the final consumer will exclusively be receiving power. A two-way network is likely to have network externalities.

ders. In other sectors like railways, there are huge technical differences in standards making it difficult to integrate services across borders. Even in the gas sector interoperability of networks is a problem due to differences in quality and other technical specifications. Urban transport and water supply are local activities which do not have many cross border relations. Thus, for some network industries, technical standardisation is a key prerequisite <sup>(1)</sup> (telecom, railways) to create an internal market and enhance European integration and the development of cross-border networks. Finally, for some network industries (transport, postal services), the frequency of supply is an important element of service quality and might be crucial to the creation of hubs (interconnection nodes). As these hubs tend to be natural monopolies and bottlenecks there is a need for creating fair access conditions while ensuring coordination of frequencies.

2.4. Country characteristics

Across countries, the same network industries are often organised very differently. In some countries huge centralised public monopolies prevail while in others the industry is dominated by private regional companies or decentralised and locally managed institutions (some-

times in public municipalities managed entities, sometimes in private/public corporations) or even in consumer managed cooperatives. Hence, network industries are largely organised according to tradition, culture and the specific conditions for the relevant sector and no apparent ‘best practice’ organisation exists. Organisation and regulation of the network industries also reflect the extent to which non-economic objectives are pursued in the form of, for instance, obligations to provide services of general interest. There are three key dimensions in the organisation of network industries:

- 1. Decentralisation/centralisation.
- 2. Vertical integration/separation.
- 3. Private/public ownership.

In France, network industries have traditionally been organised in centralised, vertically integrated public monopolies and this structure dominates (for example in EDF and GDF). In Germany, they have been organised more on *Länder* level or regional/municipality level often with narrow links between public and private companies.

Taking the electricity sector as an example of differences between Member States in Table 4 and presenting the ‘extreme’ cases gives an impression of the variety of organisational forms between Member States. The organisation and structure of network industries are

<sup>(1)</sup> In the railway sector some trains have been developed that can actually switch from one type of power supply to another and thereby overcome differences in technical standards. This, however, seems to be an exception.

Table 4  
Country differences (electricity industry as an example)

	High extreme	Low extreme
System size	Large: France, Germany, Italy, Spain, Sweden, UK	Small: Ireland (the only stand-alone network in EU), Luxembourg.
Trade	The largest exporter by far: France; Luxembourg imports 95 %.	Greece, Spain
Public/private ownership	Purely private: UK, Belgium (production).	State-owned: France.
Vertical structure	Vertically integrated: France, Italy, Greece and Ireland.	Vertically separated: Portugal, Spain, Sweden and UK.
Production methods	France is very dependent on nuclear power, Austria on hydro-power.	Denmark has no nuclear and hydro-power production power, but a relatively large share of wind power.

Source: Bergman et al. (1998) and DG II.

influenced by national cultures and traditions but also by the geographical and physical conditions in the country (size, geography, population density, natural resources, urbanisation, etc.).

The above table gives an overview of ownership structures across countries. In the UK only postal services are still publicly owned, whereas in France it is only telecom which has been (partly) privatised.

*Table 5*

**Ownership of network industries, spring 1998**

	Airlines	Electricity	Postal services	Railways	Telecom
Germany	Private	Mixed	State f	State f	Part. Private
France	State f	State	State	State	Part. Private
Italy	State	Mixed	State f	State	Part. Private
Spain	State f	Mixed	State	State	Private
Sweden	Part. Private	Mixed	State	State	State
UK	Private	Private	State f	Private	Private

*Source:* Bergman et al. (1998). A small 'f' indicates that further liberalisation is on its way.



### 3. State and process of liberalisation

#### 3.1. Current state of liberalisation by country and sector

Within Europe, liberalisation has advanced furthest in the UK and the Scandinavian countries. The UK has been in the forefront with liberalisation and privatisation starting already in the early 1980s. In the Scandinavian countries liberalisation has progressed later, but here many incumbents remain State-owned. In southern Europe, the main focus has been on privatisation rather than on liberalisation. Governments might have been tempted to sell off monopolies, because the price would be better than if competition was introduced. France and Belgium have been rather slow or reluctant to introduce competition or privatisation in the network industries.

A ranking of the five countries according to liberalisation (on the basis of the above table) would go as follows: UK, Sweden, Germany, Spain, France, and Italy.

The degree of competition and liberalisation also varies between sectors. Furthermore, there are also competitive and monopoly elements within each industry. The following table provides an example of sectors with different levels of competition.

The gas, electricity and water networks tend to be natural monopolies (see Table 7). For other sectors the monopoly is in fact open to competition, including air services and telecom for the business segment. Effective competition exists, for example, in the non-reserved postal services.

Table 6

#### Degree of liberalisation in Member States

	Germany	France	Italy	Spain	Sweden	UK
Air traffic	High	Medium	High	Medium1	High	High
Electricity	Medium1	Low1	Low1	Medium2	High1	Very high
Postal services	Medium2	Low1	Very low1	Medium1	Very High	Medium1
Rail	Low1	Low1	Very low1	Very low1	High	Very high
Telecom	High	High1	Medium1	Medium1	Very high1	Very high

Source: Bergman et al. (1998). '1' means that further measures are expected in a couple of years, '2' that these measures will be significant. For telecom the ranking has been replaced by the ranking used in Section 4 of this report.

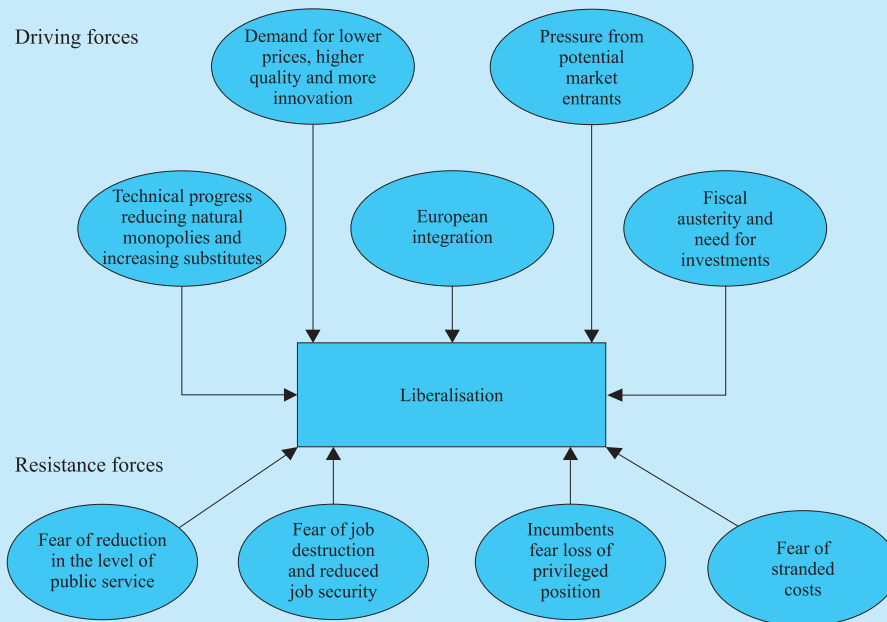
Table 7

#### Degree of competition in network industries, examples

Phase 1: Monopoly	Gas, electricity, residential telecom, railways
Phase 2: Monopoly + competition	Air services, business telecom, shipping services
Phase 3: Effective competition	Non-reserved postal services

Source: Bergman et al. (1998).

**Graph 1: Driving forces for and resistance forces towards liberalisation and competition**



### 3.2. Driving forces for liberalisation and competition

Government intervention in the network industries (in the form of subsidies, legal monopolies and other types of regulation) has often been based on the assumption of the existence of natural monopolies, externalities (network or club externalities <sup>(1)</sup>), need for income redistribution, merit goods arguments, need for security of supply, environmental externalities, etc. Government intervention also tends to aim at balancing efficiency and equity objectives. The foundation for this regulation is changing and in some cases shrinking as technological development progresses, alternative services evolve and new market-oriented policy instruments are developed. Above all, technological development has changed the *raison d'être* of traditional regulation of utilities <sup>(2)</sup>.

The importance of competition as a driving force in bringing down prices and promoting innovation has increasingly been recognised. Furthermore, new ways of introducing competition have been developed. The distinction between the natural monopoly and competitive elements of an industry has been greatly refined. As macroeconomic stability has required fiscal austerity, governments are no longer able to finance the huge investments necessary in many network industries. Hence, governments have had incentives to sell off public utilities partly or entirely and give access to private capital in the network sectors. Government budget constraints are likely to put limits to the quality of the network <sup>(3)</sup> because it is easier to postpone new investments in infrastructure than to cut down on social benefits, etc.

The driving forces behind liberalisation are thus mainly the following: pressure from the business sector and

<sup>(1)</sup> Network or club externalities imply that customers linked to a network benefit from additional customers linking up to the same network. For instance, the more people have a phone, the more people can phone each other.

<sup>(2)</sup> It is possible that liberalisation will affect the incentives to innovate and progress technologies which will again reduce the scope of the natural monopoly and facilitate a competitive market.

<sup>(3)</sup> Insufficient maintenance and new investments (for instance caused by fiscal austerity) in existing infrastructures can create excess demand and congested networks. This will spur (inefficient) by-pass and can lead to creation of uneconomic alternative infrastructures.

consumers for liberalisation because they expect lower prices and higher quality of services, pressure for fiscal austerity, pressure from potential market entrants, and pressure from a de facto undermining of the monopoly <sup>(1)</sup>.

### **3.3. Resistance towards liberalisation and competition**

Resistance to change in Europe ranges from a traditional conservative attitude through more ideologically based resistance against privatisation, free enterprise, etc. to more self-interest resistance in the form of trade unions, which foresee job losses, and the previously protected industry itself (lobbying by incumbents).

For the trade unions, the transformation to a liberalised environment also often implies a change in the status of the employees from civil servants to contract employees. This has sometimes been the core problem to tackle in a reform process. Though liberalisation reforms in the UK took place in spite of substantial resistance from the trade unions, the Scandinavian and Dutch examples have shown that Labour governments can implement liberalisation more or less in agreement with trade unions and under consideration of social objectives. In fact, liberalisation might be easier to implement under labour governments because trade unions will have more confidence in the process.

Furthermore, deregulation and liberalisation seem to have a bad name with the general public because they tend to be regarded as policies creating inequalities and leading to reductions in staff. In the present situation with high unemployment in Europe, the challenge for policymakers is to convince the general public that competition works in favour of consumers and businesses while simultaneously ensuring that job losses which might occur in a transitory period will be taken care of by training programmes, etc. but that in the long run jobs will be created because of restored competitiveness.

The public also seems to fear that the level of public service could be undermined by liberalisation. Here it is

the challenge of governments to 'market' the concept of OSGI as a way of ensuring quality, security of supply, etc. The means by which the Member States attempt to finance OSGI is open to challenge under the internal market and/or competition rules (see Section 6.3).

As many network industries involve large long-term infrastructure investments, governments often want to protect these investments by a legal monopoly. There is a risk that liberalisation will lead to falling prices, bypass, etc. which could make it difficult to recuperate the investments and result in stranded costs.

Liberalisation often involves destruction of 'economic rents' in the form of excessive price, profits and wages, low productivity or artificially high job security. The groups who benefit from such rents are often homogeneous and well organised and have a direct, clear and significant interest in avoiding liberalisation. The group that benefits from reform is much larger but more diffuse, like for instance, consumers, taxpayers, etc. This means that the benefits have to be shared by many people and the individual benefit is limited. Furthermore, the benefits will often appear in the longer run. All these factors slow down the process of reform and also imply that reforms are taking place in sectors where driving forces are larger than resistance rather than in sectors where the economic efficiency gains are the largest <sup>(2)</sup>.

### **3.4. Liberalisation and employment**

Employment problems underlie many aspects of the debate about the future of the network industries in spite of the different conditions and prospects in the industries. In most cases the immediate impact of liberalisation on employment in the specific industries is likely to be negative, but in the medium to long term the net effect could well be positive. Positive effects are expected to arise from increased demand for the products of the network industries resulting from lower prices, together with a boost to employment in other sectors resulting from lower input costs. Nevertheless, employment problems constitute an important constraint on liberalisation because of the fears of the employees of the protected monopolies that they will lose their security of employment. In addition, the short-term negative impact is concentrated on a few

<sup>(1)</sup> The legal monopoly might be undermined in practice by technological developments that create alternatives (e-mail and fax substitutions for post etc.) or by internationalisation in the sense that businesses and consumers go abroad to get their services (for instance to mail their post) or establish foreign affiliates.

<sup>(2)</sup> See Elmeskov, 1998.

sectors and is thus highly visible, while the long-term effects are more diffuse. This exacerbates the employment problems.

Although the overall benefits of liberalisation are generally accepted today, its impact on employment is difficult to evaluate and to distinguish from other factors. Increased competition accentuates the drive for higher productivity. However, quite substantial productivity improvements can be observed in most network industries even before liberalisation, possibly driven by tighter financial constraints imposed on nationalised companies by governments.

The introduction of new technologies, notably in telecommunications, seems to have been speeded up by liberalisation, but the causality is not all in one direction. New technologies increased the possibilities of bypassing the monopoly providers, thus making some of the old restrictions on competition obsolete.

In telecommunications and the energy sector at least, increased productivity has led to lower prices and a consequent increase in consumption. Long-term positive effects on employment can be expected in the telecommunication sector but, in the short term, there will be employment losses (see Section 4.5).

In the postal and water sectors, demand is much less price-elastic than in telecommunications. The net effect of productivity improvements on employment will almost certainly be negative. However, the scope for such improvements in these industries is smaller than in telecommunications and energy, because the probability of major technological innovations is small.

In railways and urban public transport, spectacular improvements in productivity can, of course, be made by reducing the network to a few heavily used routes (e.g. railways in the USA). However, major technological improvements are not to be expected and the influence of competition within each mode is probably much less important than that of competition between modes. In this respect, the first-best solution would appear to be to introduce a system of taxation and infrastructure charging which ensures that users of all modes of transport pay the full social costs. In the absence of such a system, subsidies are probably the only way of preventing a decline or stagnation of activities and employment in these sectors.

In air transport, significant productivity improvements are being achieved, partly as a result of intensified competition. However, since the market is buoyant, the loss of employment in the big national 'flag carriers' should be offset to a large extent by new entrants or the expansion of smaller airlines whose growth was previously restricted by protection measures.

In all the network industries there are frictional problems, even where there are in the medium to long term no net job losses. In addition, there is resistance to change from employees who fear that their conditions of employment and pay will deteriorate. These problems are to some extent taken into account in the formulation of Community policy, notably in fixing the deadlines for liberalisation measures so as to allow time for adjustment. Nevertheless, it is clear that liberalisation is only one of the forces affecting employment in these sectors and not necessarily the most important one.

## 4. Economic implications of liberalisation of network industries — telecommunications as an example

This section takes a brief look at liberalisation in telecommunications and its consequences for prices, quality, penetration and employment in Europe. The telecommunications sector has been chosen as it is a very profitable, capital intensive and fast growing network industry with great economic and political importance both in its own right and in terms of strategic impact on other sectors, consumers and government bodies. The sector thus sets the pace of the Information Society. The telecommunications sector is also the most advanced of the network industries in terms of liberalisation, privatisation and regulation of access pricing, etc. The sector — though not in all aspects comparable to other network industries — often plays the part of ‘role model’ for the other sectors.

Assessment of the impact of liberalisation is extremely difficult and includes the following problems:

- First, it is difficult to distinguish the impact of liberalisation from that of other factors, including technical progress (which has permitted huge productivity increases), globalisation, economic development, etc.
- Second, we do not have the necessary comparable time series for assessing the process and the impact.
- Third, in a large number of the Member States liberalisation has only just begun. It is still very early to assess the consequences. In addition, a part of the impact of liberalisation can be *ex ante*, because incumbents will react to anticipated liberalisation. Hence, many operators have reduced staff and cut prices in order to prepare themselves for liberalisation.

Furthermore, liberalisation in the telecommunications sector is a very complex process involving many different issues, like licensing, network interconnection,

operation of national regulatory authorities, allocation of telephone numbers, cost orientation of tariffs, rights of way, etc.

Because of these obstacles, this exercise can only claim to be illustrative.

The general idea is that liberalisation will lead to restructuring of the industry. This restructuring will be accompanied by changes in prices and quality of services as well as in activities and employment. Liberalisation of telecommunication markets should lead to new entrants (increased number of operators) who would establish a competitive environment resulting in a downward pressure on prices and costs. Furthermore, incentives to innovate and enhance quality should increase, leading to better products and services.

The impact on employment will be more complex. In order to be competitive, the incumbent will minimise costs and reduce staffing levels. This impact will to some extent be compensated by job creation in the new entrants, although the short run impact will probably be negative. However, as prices are falling, consumption of telecommunication services will increase (demand seems to be very price elastic) leading to job creation in the sector in the longer run. The main employment impact, nonetheless, is expected to occur in other parts of the economy as competitiveness and purchasing power increase following falling telecom prices.

### 4.1. Degree of liberalisation in the Member States

Community legislation has imposed full competition in telecommunications by 1 January 1998 in all Member States but five. These five countries (Portugal, Spain, Greece, Ireland and Luxembourg) have derogations and have postponed the introduction of full competition (until 2002 at the latest). In spite of substantial progress

of liberalisation in the telecom sector and the fact that the European markets should be fully liberalised by January 1998 (subject to the derogation) there are still many directives which are not fully implemented across the Union (see Table 9 in the Annex). In the analysis of the single market scoreboard of November 1998, the telecom sector is singled out, as the sector with the highest percentage of directives (66 %) not yet transposed across the whole Union.

The fourth Commission report on the implementation of the regulatory package <sup>(1)</sup>, which we draw on in this part, states that the bulk of the EC measures have been transposed and the report focuses on how national measures are implemented in practice. The report goes on to conclude that a considerable number of details remain to be resolved.

In order to simplify this exercise, the Member States are grouped in three broad categories (see Table 8) according to the timing of liberalisation of their telecom industries. This grouping is based on whether liberalisation happened before, at, or after 1 January 1998. This gives a simple but useful overview and limits somewhat country-specific effects. The grouping is inspired by a CEPR/SNS Report (1998) <sup>(2)</sup>.

The UK has been the leader of liberalisation of network industries in Europe. In the telecom sector, the process began in 1984. The Scandinavian countries followed in the first half of the 1990s. Germany, France and the Netherlands have respected the Community deadline of

1 January 1998. The rest of the Member States will follow later. Consequently, liberalisation has had the chance to work for some years in Group 1 and for one year in Group 2. In Group 3 liberalisation has been implemented very recently or is expected to be implemented fairly shortly.

It should be kept in mind that, as developments in telecommunications markets and regulation are very rapid, data very quickly become out of date. Therefore, the information provided here does not necessarily reflect the current state of affairs. However, for an exercise like this, where we try to provide an indication of the impact of liberalisation, this is of minor importance, because even though new developments have occurred since the time of collection of data, they still illustrate the impact that liberalisation has had in the period considered.

In the following sections, we will look at the relation between the timing of liberalisation, and its extent (indicated by number of actual operators), on one hand, and prices, quality and penetration of telecom services on the other. We expect that in liberalised markets, prices will be lower, quality higher and penetration rates greater <sup>(3)</sup>. However, other factors can affect prices, quality and penetration of telecom services. For example, one can expect that general income levels may have an effect. Geographical factors can also play a role, as can the extent of the obligations to provide services of general interest.

<sup>(1)</sup> COM(1998) 594, 'Fourth report on the implementation of the telecommunications regulatory package', November 1998.

<sup>(2)</sup> Bergman et al. (1998), 'Europe's Network Industries: Conflicting priorities', part 2, p. 153. The table presented here is slightly modified.

<sup>(3)</sup> However, it might be that liberalisation has an impact before its actual implementation because incumbents react to an anticipated liberalisation.

*Table 8*

### **Liberalisation of telecommunications in the EU**

Group 1	Competitive prior to Commission Directives	UK, Sweden, Finland, Denmark
Group 2	Fully liberalised by January 1998	Germany, France, The Netherlands
Group 3	Other	Austria, Ireland, Italy, Belgium, Spain, Portugal, Luxembourg, Greece

*Source:* Based on CEPR/SNS (1998).

However, Table 13 in the Annex shows that relatively rich countries (GDP per capita) are found in all groups (Luxembourg, Austria and Belgium in Group 3, Germany and France in Group 2, Denmark in Group 1, etc). Hence, not all the richest countries have liberalised their markets. Similarly, the UK and Finland in Group 1 belong to the middle income levels. In addition, there is a relatively high dispersion of per capita income levels within the groups. We present a simple average <sup>(1)</sup> for each country group in order to be able to compare indicators according to the degree of liberalisation.

## 4.2. Liberalisation and prices

In Graph 2 telecom prices for businesses are plotted against the number of actual operators in the Member State. Both indicators are taken from the Fourth Report

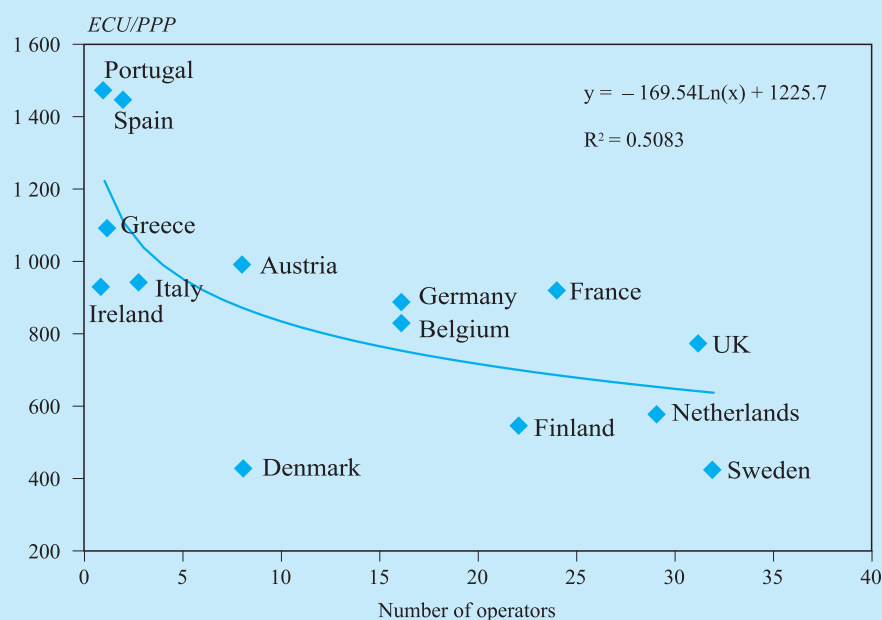
mentioned above <sup>(2)</sup>. The graph shows that prices are generally lower in countries with more operators. Member States in Groups 1 and 2 are clearly in the lower end of the price range and with a higher number of active operators.

In Graph 3 telecom prices for residential calls (consumers) are plotted against the number of actual operators. In the consumer segment, the price of identical packages of phone calls in the cheapest country (Sweden) are 37 % of the price in the most expensive country (Portugal). In the business segment, the same figure amounts to just 29 %. Generally, there seems to be a clear link between the price level and the number

<sup>(1)</sup> This means that each country has the same weight in the average because in this (political) sense each Member State has the same importance. The objective is not to find the weighted average.

<sup>(2)</sup> COM(1998) 594. As mentioned in the report (Annex p. 5) the number of operators indicates the potential competition in the market. Some operators, however, might be active only in selected segments. The figure presented here is the number of operators actually offering national public voice telephony, August 1998 (Chart 10 in the annex to the Report). Telecom prices are based on a basket of charges distributed over local, regional and long distance and over weekdays, evenings and weekends of differing duration. Luxembourg is excluded from the analysis as local calls cover the whole country.

Graph 2: Prices for business calls and number of operators





of operators. Hence, lower prices for liberalised countries can be observed both for business and for consumers (residential) <sup>(1)</sup>. The differences in the cost of business calls between liberalised countries and non-liberalised countries are larger for businesses than for consumers, indicating that businesses are likely to gain most from liberalisation.

Competition, while forcing prices to be more in line with costs, has typically an impact not only on the level of prices but also on the structure of prices. Generally, prices for long distance and international calls are reduced the most, whereas local calls and fixed charges decrease less or even increase. This phenomenon is referred to as re-balancing of tariffs. New entrants tend to target the most lucrative markets (cream-skimming) and their prices tend to fall the most. As precise figures for these segments have not been presented in the

Fourth Report we shall not look further into these aspects.

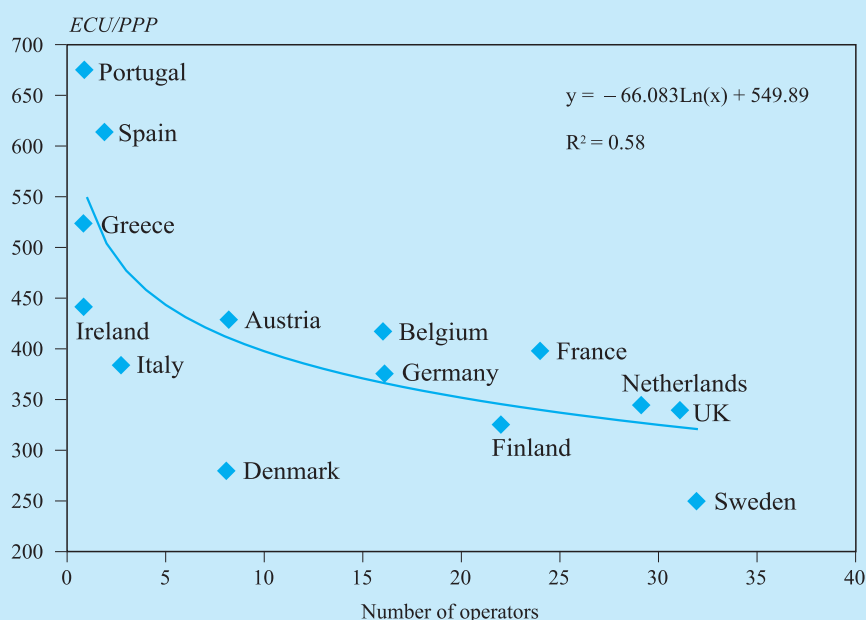
### 4.3. Liberalisation and quality

Tables 10 and 11 present some data on quality of telecommunication services. Table 10 presents data on supply time targets and actual supply times for network connection. Table 11 describes fault repair targets and actual repair times.

Quality of services is, however, very difficult to measure and compare on the basis of the available data. In addition, the information does not give the full picture of the level of quality for customers. Nonetheless, we can observe that standards for actual supply times seem to be high in Finland, France, Spain and Belgium and low in Greece, Luxembourg, Germany and the Netherlands. Based on this poor information, there seems to be no immediate link between quality in terms of supply times and the degree of liberalisation.

<sup>(1)</sup> Estimating a logarithmic function on the relation between prices and number of operators gives the following results. Business: Price = - 169xln (number of operators) + 1 225.7 with R<sup>2</sup> = 0.51. Residential: Price = - 66.1xln (number of operators) + 549.9 with R<sup>2</sup> = 0.58.

**Graph 3: Residential telecom prices and number of operators**





Turning to fault repair times as presented in Table 11, we observe an equally unclear picture. Relatively high standards seem to prevail in Austria, Luxembourg and the Netherlands, whereas Greece, Portugal and Germany have quite low quality standards. In conclusion, available data do not give a good and comparable picture of the quality of telecommunications services. The general impression, based on these poor data, is that there is no clear link between the quality of services and liberalisation.

#### **4.4. Liberalisation and penetration**

In this paragraph we discuss the impact of liberalisation on penetration. As countries with liberalised telecom markets have lower telecom prices, it should be expected that penetration in those countries would be higher. Other factors, however, might influence penetration. The extent of the obligation to provide services of general interest can have an influence as well as the income level and geographical factors.

In Table 12 penetration indicators are presented. Some of the figures, however, are from 1996 and therefore cannot be related to later developments in liberalisation. We observe that the density of main lines generally increases with the degree of liberalisation. The share of households with a phone is also higher for early liberalised countries. Luxembourg, however, has a quite high level of penetration.

For mobile phone penetration there is a clear difference in the level of the Scandinavian countries and the rest of Europe. The high penetration in Scandinavia was facilitated not only by early liberalisation but also by early implementation of a Nordic technical standard (NMT) that allowed Scandinavian manufacturers and service providers to reach critical masses at an early stage. The GSM standard, which has been established throughout Europe for quite some years now, has also been successful. However, in spite of very high growth rates throughout Europe, this has not yet led to a level of penetration of mobile telecom in the rest of Europe similar to that achieved in Scandinavia.

Geography plays a role in the costs of penetration. For instance, low density of population or difficult terrain could increase costs for constructing the networks compared to areas with easy terrain and high population density. However, in the Scandinavian countries population density is very low and topographic conditions

often difficult. As these countries show relatively low telecom prices, such geographical factors do not seem to play a significant role in determining differences in penetration between countries.

There is a somewhat higher penetration of telecom in countries in Group 1 than in Groups 2 and 3. Whether this higher penetration is a result of earlier liberalisation is difficult to say, it is generally true that countries which have liberalised their telecom markets at an early stage, have slightly higher penetration than countries which have postponed liberalisation.

#### **4.5. Impact on employment**

It is extremely difficult to measure the impact of liberalisation on employment. Employment is of course influenced by many other factors than regulation and indeed by technological progress. Furthermore, liberalisation will not only affect employment in the telecommunications sector but also obviously have effects throughout the economy. Though the impact on other sectors is likely to be dominant it is extremely difficult to estimate. In addition, the short-run and long-run impacts are likely to be quite different. As liberalisation has taken place fairly recently in most countries, the evidence of the long-term effects are of course limited.

This part discusses the results of projections made by BIPE (1997) of the consequences for employment over the period 2000–05 of rapid or slow liberalisation. The BIPE study makes a thorough inventory of the mechanisms through which liberalisation can influence employment and contains useful observations on the implications of liberalisation for the organisation of the sector, management methods, training, etc. However, it must be emphasised that the estimates have to be treated with much caution because of the limitations of the model used and the difficulty of defining the counterfactual.

According to the study, the employment impact will depend on the following factors:

- The extent to which redundancies at the incumbent will be offset by jobs in new entrants.
- The level of productivity of service providers.
- The importance of telecom manufacturing in the country.

- The impact on the rest of the economy depending on the corporate investment multiplier and the impact on household revenues.

The study suggests that liberalisation of telecommunications services will accelerate the diffusion of telecom and lead to price reductions, which will have positive effects on the economy. Lay-offs in the incumbent companies will continue throughout Europe even in countries where liberalisation began early. New operators and service providers will create employment but this will not be enough to offset the redundancies in incumbents.

The study estimates that until the year 2000 between 160 000 and 220 000 jobs will be lost in the incumbent operators in the EU as a consequence of liberalisation. This corresponds to the British experience, where BT — since liberalisation — has laid off 120 000 employees, while new entrants (including CellNet) have only created 40 000 jobs.

However, the study estimates that in the long term, liberalisation of European telecom will create or maintain 1.3 million jobs by 2005 throughout the economy. Within the telecom sector rapid liberalisation is forecast to create 93 000 jobs by 2005, whereas slower liberalisation would reduce that number substantially. While incumbents will continue to reduce their staff in both scenarios, new entrants in the sector and in other parts of the economy will create new jobs.

In all countries except Ireland <sup>(1)</sup> (see Table 14), liberalisation is expected to increase employment. Furthermore, in most countries rapid liberalisation is expected to create more employment impact than slow liberalisation. This is, however, not the case in Belgium and Italy. In these two countries, the lay-off of employees in the incumbent operator is expected to be so much larger in the case of rapid liberalisation than under slower liberalisation that job creation elsewhere will not be enough to offset this impact. Consequently, the overall positive employment impact is larger under slower liberalisation.

Nearly all job creation will take place in other sectors than telecom because the telecom sector will become a locomotive for many other sectors. Price reductions will boost purchasing power in all sectors. However, sectors taking part in the information society, exploiting electronic commerce, business consulting, research and development, etc. are likely to benefit the most. In addition, the macroeconomic stimulus which will result from this increase in demand will have further beneficial effects on the employment. The study points out that the employment impact will depend critically on the ability to restructure and adapt traditional telecom operators and the handling of renewal of skills including training policies and the social dialogue in these operators.

---

<sup>(1)</sup> In Ireland, the reduction in jobs of the incumbent will not be offset by job creation elsewhere in a slow liberalisation scenario.

## 5. Competition policy and regulation

It is sometimes argued that, if competition laws are correctly framed and efficiently enforced, specific regulation of network industries is unnecessary. A variant of this argument holds that, although sectors previously subject to legal monopolies may have to be regulated during a transitional phase, the abolition of such monopolies will rapidly lead to the emergence of contestable markets, allowing regulation to be phased out in the foreseeable future.

This view is not generally accepted, since the legacy of a century or more of legal monopoly is not likely to be overcome quickly and, in most of the network industries, a natural monopoly element will persist for the foreseeable future. Furthermore, regulation will be required as long as the industries are charged with certain public interest tasks.

If we accept the need for regulation, even temporarily, a number of questions of institutional design have to be answered. The most fundamental of these questions concern the relationship between regulation and competition policy and the level at which regulatory policies should be decided and implemented.

### 5.1. Differences between competition policy and regulation

In the network industries, the distinction between competition policy and regulation is not clear-cut. The differences are differences of degree along the following scales:

1. The sectoral scope of control: Regulatory authorities are normally sector-specific, whereas the responsibilities of competition authorities usually cover all parts of the economy <sup>(1)</sup>.
2. Objectives: The objectives of regulatory authorities normally go beyond safeguarding the conditions of competition to include, for example, consumer protection or the promotion of innovation. In some cases, the regulatory authority's duties may even seem contrary to the aims of competition policy, as when an authority is charged with protecting an incumbent firm's legal monopoly over certain services.
3. Continuity: As a general rule, regulatory authorities exercise continuous oversight over the activities of companies within their jurisdiction but competition authorities intervene only sporadically in response to complaints or specific circumstances which give rise to suspicions of anti-competitive behaviour.
4. Control over detailed management decisions: Competition policy is primarily concerned with preventing a limited set of anti-competitive behaviours and, as far as possible, ensuring a healthy market structure. Regulation usually involves detailed control over matters such as pricing and the quality of services.
5. Information requirements: Because of its more limited aims, the enforcement of competition policy normally requires less detailed information than regulation. However, some competition cases (e.g. complaints of predatory pricing) may require a thorough knowledge of the sector. In such cases, competition authorities are at a disadvantage, because the sporadic nature of their interventions makes it difficult for them to build up the necessary expertise.
6. Danger that the regulatory authority will be unduly influenced by the firms which they are supposed to control or by politicians: Undue influence by the industry over the competition author-

<sup>(1)</sup> Australia has recently established a hybrid body, the Australian Competition and Consumer Commission, which carries out the normal functions of a competition authority but also has some responsibility, shared with the states, for enforcing sector-specific legislation in the network industries. Similarly, in the UK the duties of the Competition Commission are defined not only in the general competition law but also in sector-specific legislation.

ity is less likely because of the economy-wide responsibilities of competition authorities and because they are not in continuous contact with any one sector. In most countries, the statutes establishing competition authorities ensure that they are less prone to political interference than regulatory authorities.

7. Timing of interventions: Competition authorities normally intervene after the fact, although in some cases prior authorisation may be required (e.g. mergers under the EC Merger Regulation). Regulatory authorities intervene at all stages and, in particular, they make specific orders governing the future conduct (e.g. pricing) of the regulated companies. The procedures of competition authorities also tend to be more cumbersome and time-consuming than those of regulatory authorities.

## 5.2. The need for regulation

Before deciding whether there is a need for regulation — or, indeed, for competition policy — it is necessary to define the aims of public policy in relation to the sectors in question. We take the main objectives of public policy in relation to network industries to be the following:

1. To introduce competition wherever this would enhance welfare.
2. To protect customers and suppliers from abuse of dominant positions.
3. To achieve optimal use of costly infrastructure.
4. To ensure that there are adequate incentives to invest, reduce costs, improve quality and innovate.
5. To prevent serious disruptions of service or supply.
6. To ensure the provision of defined levels of certain services at 'affordable' prices (variously known as 'services of general interest', 'public service obligations' and 'universal service') — see Section 6.
7. To ensure an adequate level of consumer protection (e.g. complaints handling and dispute settlement).

In a market where 'workable competition' can be ensured, it is generally agreed that regulation is unnecessary for the achievement of objectives 2, 3 and 4 if an adequate competition policy is enforced. However, the network industries present three features, which suggest those competition policy instruments alone, would not be adequate.

The first of these special characteristics is the existence (except perhaps in the postal sector) of a 'bottleneck' infrastructure which constitutes a natural monopoly. In most of the network industries, it is not feasible to remedy this problem by duplicating the infrastructure, not only because of high investment costs but also because of important environmental externalities. In telecommunications, however, the importance of the bottleneck infrastructure (the local loop) is diminishing with the development of mobile telephony and an alternative fixed infrastructure in the form of cable networks. The efficiency of the network industries depends on the existence of an adequate mechanism for controlling abuses of monopoly power over access to this infrastructure. The main problems associated with access conditions are discussed in Section 5.5.

The second distinguishing feature of the network industries is the fact that, until recently, they were monopolies not only at the level of infrastructure but also at the level of final services. The former monopolies are still the dominant firms in their markets and have at their disposal a wide range of instruments to counter the threat of new entrants.

The third special feature is that the efficient provision of network services usually requires a high degree of coordination between operators. In the transport sector, for example, travellers need a central point of access to information about all available services and the shortest possible waiting times between connecting services. In the electricity and gas industries, the flows through the network must be centrally controlled. This need for coordination can conflict with the objective of opening up the markets to competition if it leads to collusion.

A study carried out by the CEPR and the SNS (see Bergman et al., 1998) suggests a stylised evolution of the network industries in three phases. The first phase corresponds to the old structure of legal monopolies, where the public interest was normally deemed to be safeguarded by public ownership. In principle, the State could intervene directly in management decisions to

protect consumers from abuse of monopoly power. The industries now find themselves at various points in the second phase, where the old monopolies are being broken up and a new system of regulation is being established. In this phase an independent regulatory body is usually established. The tasks of this body include not only controlling the quantity, quality and pricing of final outputs but also preventing the incumbent from creating artificial barriers to new entry. It therefore becomes important to 'unbundle' the services provided by the incumbent and to control the supply and pricing of intermediates, such as access to essential facilities. At some point during the second phase, the gradual introduction of competition should make it possible to reduce the intensity of regulatory control. The third phase would begin when competition is extensive and the need to regulate monopoly power diminishes, perhaps to the point where only the normal competition rules are necessary.

However, it seems unlikely that sector-specific regulation will become entirely redundant in the foreseeable future in any of the sectors, with the possible exception of telecommunications. Detailed control over the conditions of access to bottleneck infrastructure will continue to be necessary unless technological developments eliminate the natural monopoly. Sector-specific regulation — but not necessarily a sector-specific authority — will be essential for this purpose.

It can also be argued that the continuing dominant position of the old monopoly-holders requires specific regulatory measures because of the cumbersome and time-consuming procedures required for the enforcement of competition law relating to abuse of a dominant position. Furthermore, the technical complexities of the network industries seem to require a specialist staff charged with continuous oversight over the activities of the firms in question.

Finally, it seems clear that where governments impose obligations on operators in the network industries to provide services of general interest, there must be some form of regulation to ensure that the obligations are fulfilled. This is also true when the obligations are the subject of contracts freely entered into by service providers in return for compensation, as in the air transport sector.

### **5.3. The relationship between regulation and competition policy**

If sector-specific regulation is a continuing necessity, how should responsibilities be divided between competition authorities and regulators and how should their activities be coordinated?

In most cases, the responsibilities of competition and regulatory authorities overlap to a large extent. In communications concerning the application of competition rules to the telecommunications sector <sup>(1)</sup>, the Commission clearly takes the view that nothing in the sector-specific rules should prevent the application of normal competition rules but that the former should prevail where they impose a tighter control over the dominant firms. This implies the following division of responsibilities:

- Competition and single market law provides the general framework within which regulation operates. Sector-specific rules may add to this framework but should not normally detract from it.
- The regulator is responsible for applying the sector-specific rules, which will often obviate the need for intervention by the competition authority.
- The competition authority (or the courts) intervenes when the regulator does not have power to ensure that the horizontal competition rules are respected, when (s)he fails to take action to enforce those rules or when (s)he acts in contradiction to those rules.

This division of responsibilities, which respects the legal convention that specific laws prevail over general laws, has many advantages. It gives regulators, with their specialised knowledge and more rapid and flexible procedures, discretion to act within an overall framework set by competition law, while reserving the more cumbersome procedures of competition law for exceptional cases. Regulation can satisfy the need, in sectors undergoing rapid change, for measures to prevent the abuse of dominant positions, whereas competition law (Article 86 EC) can only be applied retrospectively.

<sup>(1)</sup> 'Guidelines on the application of EEC competition rules in the telecommunications sector' (OJ C 233, 6.9.1991) and 'Notice on the application of competition rules to access agreements in the telecommunications sector' (1998).



Furthermore, this system provides a means whereby the Commission and the ECJ in matters, which affect the integration of the EU (see Section 5.4), can combine decentralised regulation with centralised oversight.

In some cases, however, sector-specific legislation allows exemptions from the normal competition rules. For example, the Postal Directive allows Member States to grant a legal monopoly over 'reserved services' as compensation for obligations to provide services of general interest. Exemptions should be justified by peculiarities of the sector concerned and their scope should be clearly delimited in order to avoid conflicts between regulatory and competition authorities.

#### **5.4. Subsidiarity and the level of regulation**

The principle of subsidiarity requires answers to the following questions:

- At what geographical level should the regulatory framework be decided?
- Should the framework be imposed from above or established by voluntary agreement between national or regional governments?
- At what level should the rules be implemented?
- If implementation is decentralised, to what extent should there be central oversight or coordination?

It is probable that the answers to these questions will differ not only according to the industry concerned and the phase of liberalisation, which has been reached but also according to the nature of the problems addressed by specific rules. The existence of important cross-border externalities would justify the establishment of rules at Community or international levels. Thus, a broad geographical coverage is more clearly indicated for telecommunications, for example, than for local public transport or water. However, even in a sector like telecommunications, the EU does not have a clear interest in all regulatory matters. EU intervention in the field of access to infrastructure seems to be clearly justified, since the functioning of the single market can be affected. The extent of the EU's interest in matters such as consumer protection or public service obligations needs to be further analysed in the light of the provisions of the new Article 16 of the EC Treaty. Even in fields where the EU does not exercise its legislative

powers it may have an important role to play in organising cross-border consultations and the exchange of information and experience.

As far as implementation is concerned, there are strong arguments in favour of decentralisation in terms of better information and speedier decision-making. Where there is no consensus about the correct solution to a particular problem, decentralisation also permits experimentation with a range of options. One disadvantage of decentralisation is that it strengthens the influence exerted by the regulated industry, since decentralised regulators are in constant contact with a limited number of operators and only one national government. A supranational regulator, being subject to more diverse pressures, should find it easier to maintain his/her independence.

Regulatory authorities must be allowed some discretion in implementing policy and it seems improbable that legislation can be framed in such a way as to exclude the possibility that the exercise of this discretion could affect economic integration. In the EU context, it therefore seems inevitable that the Commission, as the Community's competition authority also charged with ensuring that single market legislation is respected, will be called upon from time to time to review the decisions of regulatory authorities.

A system of coordination of national regulators at the European level may offer many advantages. While retaining the advantages of decentralisation, such as respect for the principle of subsidiarity and better informed regulatory authorities, coordination would provide a forum for resolving cross-border problems by common agreement, for exchanging information and experience and establishing 'best practice' guidelines.

#### **5.5. Access to the bottleneck infrastructure**

The question of conditions of access is fraught with difficulties, particularly in sectors where the operator of the bottleneck infrastructure (essential facility) is also a provider of final services. Regulators have to attempt to satisfy a number of different requirements:

- The operator must have an adequate return on its investment and an incentive to expand and improve the network.

- If effective competition is to develop in the downstream market, the conditions of access must be non-discriminatory and the access price must not be so high as to discourage entry.
- If alternative infrastructures are available or being developed (as in telecommunications), the infrastructure operator must not be allowed to practice predatory pricing in order to exclude competitors.
- The access price should not be so high as to encourage inefficient bypass (i.e. the use of roundabout routes) or the construction of duplicate networks when the existing network is under-used.

Optimal use of the infrastructure could be achieved if the access price were set at marginal cost. However, the fixed costs of network infrastructure are normally very high in comparison to the variable costs. Marginal cost pricing would therefore result in a substantial loss for the infrastructure operator. One solution to this problem would be to provide State subsidies to cover the fixed costs. The disadvantages of this approach are the high cost to the taxpayer and the risk that investment decisions would be determined by government rather than in response to demand from the infrastructure users.

Many other methods have been proposed for controlling access prices. A consensus seems to have emerged in favour of some form of cost-based pricing rule, although regulators need very detailed and accurate information on costs in order to enforce such a rule. Furthermore, cost-based pricing is not optimal in situations where prices diverge from marginal costs, i.e. where there are large fixed costs. In such situations, a mark-up above marginal cost is needed to recover the fixed costs. Usage-based pricing, which varies the mark-up according to the demand in the different downstream markets, is a more efficient way of recovering fixed costs than cost-based pricing.

Regulatory control of usage-based pricing is relatively easy in sectors where the bottleneck infrastructure is provided by a company which is not present in the downstream markets. In such circumstances, a global cap on access prices is sufficient, as the infrastructure operator has no incentive to use price discrimination as a means of excluding service providers from the downstream market. However, if the infrastructure operator also competes in the downstream markets, individual

access prices have to be controlled. The individual control of usage-based prices would present enormous difficulties because it would require a thorough knowledge not only of the operator's costs but also of demand conditions in all the downstream markets.

When the operator of the bottleneck infrastructure is also present in the downstream market as a service provider (as in telecommunications), the task of the regulator is particularly difficult. In this case, the regulator needs to ensure that the accounts of the infrastructure and service provision activities are separate and that the operator's access conditions do not discriminate against other service providers. Since, in these circumstances, the infrastructure operator is also likely to be the dominant firm in the market for final services, the regulator will probably have to control not only the access prices but also the prices charged to consumers. The special danger in this situation is that the operator will apply a 'price squeeze' on entrants, by charging high access prices and low prices for final services.

## **5.6. Competition in the market and competition for the market**

In the network industries there are a number of markets which cannot be opened to full competition. The provision of bottleneck infrastructure is the most extreme case. Other markets which are difficult to open up include the provision of unprofitable services of general interest (see Section 6) and the provision of certain services when infrastructure capacity is tightly constrained. Examples of the latter type of market are to be found in the railways and air transport sectors, where the possession of a 'slot' at a peak time can give an operator considerable market power when only a small number of slots is available.

One way of dealing with the problem of inadequate competition in the market is to institute competition for the market, by awarding the right to operate the infrastructure or services in question to the firm which makes the most advantageous offer in a tendering procedure. The criteria for selecting franchisees could include the price offered for the franchise or the amount of subsidy demanded and undertakings by the bidders concerning the quality and price of services. This approach has been adopted at the EU level for allocating public service obligations to airlines under Council Regulation (EEC) No 2408/92, in the UK for granting

franchises to railway operators and in Denmark for regional and local bus services. The auctioning of franchises poses fewer problems when applied to the operation of services than when applied to the infrastructure. Where the operation of infrastructure is concerned, the

maintenance and development of the network are key tasks. It is difficult to devise a franchise contract, which provides adequate incentives for an operator to carry out these tasks in the face of uncertainty about whether the franchise will be renewed.



## 6. Services of general interest and the operation of the single market

Some network industries produce goods or services corresponding to very basic needs of the population. Therefore, it is widely politically accepted that all consumers should have access to a defined set of services on reasonable conditions. This is the reason why OSGI (obligations to provide services of general interest) occupy a prominent place in the policy debate regarding these network industries in the EU. Various issues are at stake: how to define OSGI, what is their rationale, how to finance them, how to reconcile competition and OSGI, how liberalisation may lead to an evolution of this concept. These issues are discussed below.

### 6.1. Definition of OSGI

Obligations to provide services of general interest may take various forms: provision of a minimum service or of a minimum quality at a reasonable or affordable price, tariff balancing such as a uniform price across regions or across consumers groups, in spite of differences in the cost of providing the service. The objective of these obligations is to make sure that everyone has access to a defined set of services at prices they can afford.

In a communication (COM(96) 443 of 11 September 1996) the Commission has described its view on services of general economic interest and underlined some political principles. The Commission has emphasised the freedom of Member States to define OSGI and how they should be regulated, administered and financed. Furthermore, the Commission has stated — with reference to Article 222 of the EC Treaty, that it is neutral as to the question of ownership of the enterprises providing these services (whether private or public) and that competition questions will be regarded independently of this status.

In the following the concept of OSGI is used in its widest sense covering all specific obligations related to

the industries in question whatever form they might take. In some sectors, telecommunications for example, the more narrow and well-defined concept of USO (universal service obligations) is used, whereas the concept of public service obligation is used in the transport sectors.

The difficulty with the concept of OSGI is that it leaves open for interpretation what constitutes the defined set of services or 'affordable prices'. A precise definition of the OSGI is, to a large extent, country and sector specific. For example, for postal services, basic services can be defined as access to specified letter and parcel services of a certain quality with respect to time and reliability of delivery. For telecommunications, they can correspond to voice telephony, emergency services and call boxes (quality with respect to waiting time for connection, the frequency of faults, repair times, etc.). The concept of affordability is also closely linked to national circumstances. Therefore it cannot be defined at the Community level and Member States tend to apply it in a non-measurable manner.

In the absence of EU regulation, different national regulations could create competition problems, leading to entry barriers and hindering the proper operation of the single market (see Sections 6.4 and 6.5). This means that guidance should be given to Member States, as in the case of the telecommunications industry, where the Commission has issued a directive.

### 6.2. *Raison d'être* of OSGI

OSGI may be the outcome of a political process or result from regulatory capture by some pressure groups <sup>(1)</sup> and they are not necessarily motivated by social welfare

<sup>(1)</sup> For example, entrants successfully lobby in favour of strict restrictions on the historic operator's pricing policy with the intent of weakening its competitive position or the historic operator may favour the maintenance of a stringent USO as this may justify some of its privileges.

considerations. However, several arguments can be given to justify OSGI: network externalities, provision of public goods, regional policy and redistribution concerns. But it is very difficult to judge whether an OSGI is well founded or due to regulatory capture.

OSGI make it possible to correct market inefficiencies in the presence of network externalities. Network externalities arise when the benefits from using a network depend on the number of individuals who are connected to the network. For example, in a telecommunication network, the number of subscribers determines the number of individuals with whom any particular user can communicate. Such externalities may lead to an inefficient outcome (for example, participation too low) in an unregulated market and these inefficiencies may be reduced through regulatory measures promoting the affordability of access to the network. However, this argument has some limitations. First, it does not apply to all the network industries where an OSGI is imposed (e.g. in the electricity, gas or water sectors). Second, some operators may benefit from the exploitation of network externalities (they increase the consumer's willingness to pay) so that there is no need for regulatory measures. The study of the University of Toulouse (1997) also downplays the role of network externalities.

Another argument is that the OSGI is a mechanism for the provision of merit goods. The idea is that the existence of a communication or transportation network is valuable in itself because it 'binds the nation together' and is essential for the functioning of an economy. Similarly, energy and water are basic necessities. Therefore, the diverse benefits from these services should not be distributed purely on the basis of ability to pay.

The OSGI can also be an instrument of regional policies. For instance, uniform pricing may be a way to subsidise rural customers in order to encourage households and firms to locate in rural areas. Similarly, maintaining basic public services (like post offices) in small villages may contribute towards preventing the decline of rural areas. However, the relationship between universal access to some networks and regional development is complex and overall positive impact is not always guaranteed. For example, the access to an efficient transportation network may speed up the decline of a region instead of fostering its development.

Finally, OSGI may be justified by a redistribution concern. OSGI makes it possible to organise redistribution between categories of consumers: to the high-cost consumers and to the low income ones. However, it can be more efficient to provide the poor with income rather than to encourage consumption of specific goods and services. Therefore, even if redistribution through prices has good foundations, this is not necessarily the best means of achieving this objective and one should make a cost/benefit analysis of alternative redistributive policies.

### **6.3. The provision and financing of services of general interest**

The costs of OSGI <sup>(1)</sup> can be defined as the net loss incurred by an efficient operator in providing these services. As pointed out in the CEPR/SNS study (Bergman et al., 1998), there are three broad categories of costs associated with OSGI: the costs associated with serving uneconomic areas such as remote regions, mountainous areas and areas where population density is low, the costs associated with uneconomic customers such as low income households and unemployed people and the costs of uneconomic services which are non-profitable services such as the provision of public call boxes. To minimise the costs <sup>(2)</sup> of providing services of general interest, it is important to find mechanisms that provide utilities with the right incentives.

In this respect, the issue of the provision of services of general interest has to be distinguished from the issue of its financing. Regarding the provision of these services, the study of the University of Toulouse (1997) argues that an attractive method is to organise an auction in which firms announce the compensation they demand to fulfil a specific obligation in a given area and the

<sup>(1)</sup> See for instance, COM(96) 608: Communication on the costing and financing of universal service in Telecommunications or Annex III in the Interconnection Directive (97/33/EC).

<sup>(2)</sup> According to the CEPS report, two concepts of costs are widely used: fully distributed costs and avoidable costs. Following the first concept, all costs including overhead and joint costs are distributed among the services (OSGI and non-OSGI) according to some convention. The choice of the convention is arbitrary and will, when the services of general interest are subsidised, give the utility an incentive to allocate as many common costs as possible to such services. The avoidable costs are the costs that could be avoided if the services in question were not provided. Avoidable costs do not include the cost of equipment that is needed to provide the services of general interest but which is also needed for other services and would therefore have to be supplied and pay anyway. The assumption is that OSGI have no influence on the decisions concerning the production apparatus. This is a correct analysis in most cases but not always.

lowest bidder is selected. The main advantages of this method are that if the auction is truly competitive, it tends to ensure that the OSGI is assumed by the most efficient operator at a close to minimum cost and that it permits to increase the transparency on the costs of OSGI. But this auction scheme will not work successfully if the number of expected bidders is small and if collusion amongst bidders cannot be ruled out (see Section 5.6). The use of this method for the provision of OSGI is also consistent with different ways of financing it.

Regarding the financing of OSGI, it is not possible to define a single mechanism that would be appropriate in all sectors and in all countries. Indeed, the appropriate financing mechanisms will involve various trade-offs which are to a large extent sector and country-specific (for example, competitive neutrality of the mechanism implies no excessive protection of the OSGI operator, which would deter entry, but appropriate compensation for the burden of OSGI is necessary to avoid phenomena such as 'cream-skimming').

Therefore, the costs of OSGI may be financed in a variety of ways: through internal cross-subsidies between an incumbent's product prices, through access charges, through compensatory payments from a universal service fund financed by all firms in the industry and through direct government subsidies. A limited number of Member States (Italy, France, Ireland and Spain) have put in place an active universal service funding system for telecommunications, whereas others consider that the costs involved are too low to justify such mechanisms.

However, liberalisation implies that the traditional method of funding services of general interest, i.e. cross-subsidisation, is not sustainable for two reasons. First, if new entrants put downward pressures on prices, in particular in those markets serving higher value customers, this cream-skimming process will undermine the incumbent's ability to finance services of general interest through cross-subsidisation. This has in some cases led to the establishment of a reserved area (a legal monopoly) for an incumbent in order to avoid cream-skimming. Second, where an incumbent sets prices above cost in order to OSGI, inefficient entry may be stimulated. In addition, this method has hidden the costs of providing these services and has thus not given utilities the right incentives to minimise these costs.

The method based on access charges consists in taxing all firms using the incumbent's essential facilities. This approach also raises several difficulties. By forcing access charges to lie above costs, it can lead entrants to unnecessarily duplicate parts of the incumbent's facilities in an attempt to bypass the relatively expensive facilities it provides. By adding a levy onto access charges, an incumbent is also in a position to increase its rivals' costs, creating entry deterrence and foreclosure. Cost-based pricing rules can reduce these disadvantages but are difficult to apply, due to heavy information requirements.

OSGI can also be financed through a universal service fund to which all firms of the industry contribute. This method, which has been considered in some liberalised telecommunications markets, is more transparent and does not provide incentives to inefficient bypass. Problems with this approach are connected with the estimate of the incumbent's costs of universal service, the allocation of these costs across the firms in the industry and the fact that it reinforces the position of the incumbent by making it a special player in the industry. In addition, fund contributions might constitute an entry barrier. A way to estimate and minimise costs could be to introduce the system of auction defined above. In economic terms, the least distorting way of financing OSGI is through direct subsidies. An appropriate solution could then be to combine the system of auction with direct subsidies.

## **6.4. OSGI and competition**

### *6.4.1. The Treaty provision*

The new Article 16 of the EC Treaty requires the Community and the Member States to take care that services of general economic interest 'operate on the basis of principles and conditions which enable them to fulfil their missions'. However, neither this article, nor Article 90, defines what is meant by service of general economic interest and their missions. The Commission's Communication on 'Services of general interest in Europe' <sup>(1)</sup> also stops short of providing any generally applicable operational definition. Notions of general interest, public service and universal service have developed in a piecemeal fashion and, as already mentioned, differ between sectors and Member States.

<sup>(1)</sup> COM(96) 443 final, 11.9.1996.

Until now, the basic approach of Community policy has been to try to establish whether and to what extent exclusive rights, subsidies or special financing arrangements are necessary for the fulfilment of the OSGI set by the Member States. But, in principle, competition cannot be distorted on the pretext of fulfilling these obligations.

The Treaty in Article 90(1) explicitly allows the Member States to grant special rights and obligations to certain companies but makes it clear that any measures taken by Member States have to comply with the general competition provisions. However, the second paragraph of this article allows for the possibility that undertakings operating services of general economic interest may be exempted from the competition rules in the Treaty: 'Undertakings entrusted with the operation of general economic interest ... shall be subjected to the rules ... on competition, in so far as the application of such rules does not obstruct the performance, in law or in fact, of the particular tasks assigned to them.' Finally, Article 90(3) entrusts the Commission with the implementation of these rules, possibly using directives and decisions.

The provisions of Article 90 establish two important principles. First, the legitimacy of the general economic interest motivation for regulation is recognised. Second, the primacy of competition rules is established in principle but general economic interest may override competition principles in cases of conflicts. This might hinder the process of liberalisation in some industries. For example, incumbents might have an incentive to challenge domestic regulatory provisions in court, attempting thereby to reduce the scope of services opened to competition.

#### *6.4.2. Is competition consistent with OSGI?*

When services of general interest have been selected, the organisation and conditions of their provision has to be decided. OSGI are often used as an argument for monopoly and against competition. As these services may imply a financial loss for the service provider, they must be subsidised. The traditional way of solving this financial problem has been via cross-subsidisation from profitable services to loss-making services. To protect the incumbent utility against competitors only interested in offering the profitable services (cream-skimming), the former is provided with exclusive rights.

However, there are arguments that show that this traditional way of financing universal services is not the most efficient (see above, Section 6.3) and that competition can contribute towards the development of services of general interest.

First, it is not always the case that the provision of services of general interest is a loss-making activity. For example, network industries with large infrastructure costs will have an economic interest in achieving as high coverage as possible within their area. It may also be profitable for a firm in a network industry to set relatively low subscription charges to stimulate participation and set relatively high usage charges to recoup fixed costs. This form of pricing explains the rapid growth in subscribers on mobile telephony networks in Europe. Second, in spite of exclusive rights, public utilities are not always able to provide services of general interest in good conditions. For example, the long waiting times for a telephone connection in some Member States is a clear illustration of this phenomenon. Finally, the utility tradition of average cost pricing will sometimes exclude consumers with low incomes and consumption and exclusive rights and cross-subsidisation are far from being always the most efficient solution for the provision of services of general interest.

The introduction of competition can in some cases contribute towards the development of services of general interest. For example, in the telecommunications industry, competition is likely to speed up the introduction of new technology and thereby price decreases, thus contributing to 'general interest' objectives. Market power can act against these objectives because firms might tend to set excessively high prices, leading to lower subscriber numbers and reducing network externalities. But although competition can help, the existence of high cost or low income consumers and loss-making services means that some regulation is still needed to attain 'general interest' objectives. The division of responsibilities between regulatory and competition authorities has been discussed in Section 5.3.

### **6.5. OSGI and the single market**

Within the framework of the single market, the European Commission accelerated its programme of liberalisation and harmonisation in the network industries. Liberalisation means establishing the conditions of market rules, necessary for the operation of an internal market and harmonisation aims at bringing consis-

tency across industries predominantly shaped by national markets.

In a competitive international environment, it is extremely important that services of general interest are clearly defined and the conditions of their provision are transparent so as to avoid compromising the benefits flowing from an internal market.

The operation of the single market can be endangered for two major reasons. First, where entrants are required to contribute towards OSGI, this can create a barrier to entry even if the contributions are calculated in a fair and objective way. Such contributions constitute a cost outside the control of the entrant. If they represent a substantial proportion of the entrant's total costs, they can significantly reduce its room for manoeuvre in price competition with the incumbent. Second, differences in the definition of OSGI between Member States may

create different conditions of competition, bearing on the international competitiveness of national firms and creating trade distortions.

In sectors with local (water, urban transport,) or even national networks, the national definition of OSGI should normally not create distortions. However, in sectors where networks develop across national boundaries (like telecommunications), it is necessary to avoid competition between regulators and discrimination between national and foreign firms. These considerations plead in favour of some harmonisation of the concept of services of general interest but, as the tradition of the different Member States is very different in this area, the political and legal obstacles to such harmonisation would be formidable. Nevertheless, there is a role for central coordination to avoid obstacles to economic integration and this role might be played by the Commission.

## 7. Lessons from the Community's perspective

The European Union has harmonised basic regulation in the network industries and gradually introduced competition in order to set the scene for an internal market in network industries. But liberalisation and indeed privatisation remain mainly Member State issues. Furthermore, because of huge discrepancies between countries, the European Union has to concern itself with fairly general terms (guidelines, communications, and framework directives and process facilitator) and leave specific regulation to the Member States. Similarly, because of significant sector-specific characteristics, it is difficult to consider these industries within an overall generic framework. Therefore, the design and geographical scope of the regulatory framework will differ according to the industry. For example, in those sectors where there are important cross-border externalities such as air transport, there is a case for rules at the Community or international levels while in others such as local public transport, different national regulations should not create problems.

This does not mean that there is no specific role for the European Union. First, the Union has to ensure that the internal market functions well, while taking account of the objectives of social and economic cohesion. Second, there is certainly an important role of central coordination and of delivery of information to be played at the Union level. More precisely, this aspect should be evolving along the following lines:

- Facilitate liberalisation of national network industries: present arguments and evidence, develop

policies, coordinate viewpoints, and so on (Green Papers, reports, debate, etc.).

- Promote and facilitate European integration of network industries in order to improve the functioning of the internal market and issue general rules harmonising the legal environment (TENs, SMP, etc.).
- Facilitate fair and efficient competition in network industries through competition policies (merger control, state aid, etc.).
- Coordinate regulatory bodies (directives, guidelines, and institutions) and ensure that their enforcement power (and their ability to act independently of both government and operators) is similar in all the Member States.

Finally, it is important to point out that not only should the institutional design vary by industry but it should also evolve over time. As we have seen in Section 5, the nature and the importance of regulations vary according to the degree of liberalisation already achieved. Therefore, it is particularly important that the performance of regulatory agencies is periodically evaluated and that the possibility of abolishing a regulatory agency is recognised. The Union can also play a role in this assessment in order to ensure that similar principles are applied in the different Member States. In this respect, it would be useful to ensure similar levels of enforcement power by the different national regulatory agencies.



## 8. Conclusion

This report has shown that network industries like telecommunications, energy (electricity, gas), transport (air, local public transport, rail), water and postal services which have been traditionally sheltered from competition and operated within national or local boundaries have experienced important changes in all Member States over recent years. These network industries make up 5 % of European GDP and employment. However, their economic importance is larger than that indicated by this ratio as the price and quality of their outputs are important for the growth and competitiveness of European industries, for the well-functioning of the internal market and for the standard of living of European consumers.

Despite a common trend of liberalisation, significant differences exist between sectors and across countries. The degree of competition is higher in the telecommunications industry and in the non-reserved postal services and lower (often non-existent) in the water and railways industries. Between these extremes, one finds the energy sectors, the other transport services and the postal services. These industries are also organised very differently across countries. In some countries, network industries have traditionally been organised in centralised, vertically-integrated public monopolies (France), while in others, private regional or local companies play a significant role (Germany). Similarly, the liberalisation process is most advanced in the United Kingdom and the Nordic countries and least advanced in southern Europe. These differences explain why it is difficult to analyse these industries within an overall generic framework.

The liberalisation process in network industries is very gradual in Europe because there are simultaneously driving forces in favour of this process and elements of resistance against it. The main driving forces are technological developments, the constraints on public finances, the European integration process and the recognition of the importance of competition as a factor

bringing down prices and increasing innovation. The main forces of resistance are the lobbying of incumbents, the resistance of employees and trade unions because of the risk of job destruction and equity considerations making it necessary to ensure access to essential services at affordable prices.

Even if it is extremely difficult to measure the economic impact of this liberalisation process, one can expect efficiency gains in the medium to long term, leading to lower prices, increased demand and competitiveness improvements. However, the changes brought about by competition also imply adjustment costs in the short term. These adjustment costs include short-term employment losses, even if part of these losses is due to technological developments. Shifts in technology, ownership and market structure also create big regulatory and organisational challenges as the nature of regulation and the kind of organisation have to evolve over time with the degree of liberalisation and need to take into account objectives of social welfare such as those pursued by obligations to provide services of general interest.

More precisely, there is an increased role for competition in these industries and therefore a need for a great coordination between regulatory and competition authorities. The division of responsibilities is that competition and single market law provides the general framework within which regulation operates while the regulators, with their specific knowledge and more rapid and flexible procedures, are responsible for applying the sector-specific rules, within the framework set by competition law. The design and the geographical level of the regulatory framework will also differ according to the industry and the nature of the problems addressed by specific rules. For example, the existence of important cross-border externalities, such as in telecommunications or air traffic control, would justify the existence of rules at the Community or international levels. This institutional design and the respective role

of regulatory and competition authorities must evolve over time with the degree of liberalisation reached by a particular network industry. It is therefore important to ensure that the performance of regulatory agencies is periodically assessed.

Obligations to provide services of general interest (OSGI) occupy a prominent place in the debate regarding the liberalisation of network industries in the EU. The main objective of these obligations is to make sure that everyone has access to a defined set of services at prices they can afford. OSGI are often used as an argument against competition. However, the introduction of competition can also speed up the introduction of technology and thereby lead to price decreases, which is beneficial for universal service. Although competition can help in attaining universal service, the existence of high cost or low-income consumers implies that some regulation is still needed to attain universal service objectives.

The costs of OSGI may be financed in a variety of ways: through internal cross-subsidies between an incumbent's product prices, through access charges, through compensatory payments from a universal service fund financed by all firms in the industry or through direct government subsidies. However, financing through a universal service fund, which has been considered in some liberalised telecommunications markets, offers advantages: it is transparent and does not provide an incentive to inefficient bypass. Problems with this approach are connected with the estimation of

the costs of providing services of general interest, the allocation of these costs across the firms in the industry and the fact that it reinforces the position of the incumbent by making it a special player in the industry. A way to estimate and minimise costs could be to introduce a system of auction for the provision of universal service. In economic terms, the least distorting way of financing OSGI is through direct government subsidies. An appropriate solution could then be to combine the system of auction with direct subsidies.

A precise definition of the OSGI is, to a large extent, country and sector specific and until now, this is a matter for the Member States. But different interpretations in different Member States could create competition problems and hinder the operation of the single market. For example, differences in the definition of OSGI may lead to different conditions of competition and distort trade. In particular, in sectors where networks develop across national boundaries, such as telecommunications, it is necessary to avoid competition between regulators and discrimination between national and foreign firms.

Therefore, from the Union's perspective, the major challenges in relation to network industries are to define good principles of guidance facilitating liberalisation while not jeopardising the development and proper functioning of the internal market. It is also important to coordinate national regulatory bodies, to contribute to a periodic assessment of their tasks and to bring about greater consistency in regulatory enforcement across the EU.



# Bibliography

Bergman, L. et al.: 'Europe's network industries: conflicting priorities — telecommunications', *Monitoring European Deregulation No 1*, Centre for Economic Policy Research (CEPR) and Centre for Business and Policy Studies (SNS), 1998.

BIPE Conseil: 'Effects on employment of the liberalisation of the telecommunications sector', study prepared for the European Commission, January 1997.

CEPS: 'Towards a single market in utilities', report of the CEPS Working Party on Utilities, July 1996.

EITO: European Information Technology Observatory 1998.

Elmeskov J., 'International trends in structural policies', *Danish National Economic Journal* (in Danish), OECD, No 1, 1998.

EuroData: 'T-basket', telecommunications cost indicators using the OECD basket methodology, May 1998.

European Commission, 'Guidelines on the application of EEC competition rules in the telecommunications sector', OJ C 233, 6.9.1991.

European Commission, 'Services of general interest in Europe', COM(96) 443 final of 11.9.1996.

European Commission, 'First monitoring report on Universal Service in Telecommunications in the European Union', COM(1998) 101.

European Commission, 'Third report on the implementation of the telecommunications regulatory package', COM(1998) 80, 18.2.1998.

European Commission, 'Fourth report on the implementation of the telecommunications regulatory package', COM(1998) 594, 25.11.1998.

European Commission, 'Universal Service for Telecommunications in the perspective of a fully liberalised environment', COM(1996) 73, 13.3.1996.

European Commission, 'Assessment criteria for national schemes for the costing and financing of universal service in telecommunications and guidelines for the Member States on the operation of such schemes', COM(1996) 608, 27.11.1996.

European Commission, 'Notice on the application of competition rules to access agreements in the telecommunications sector', OJ C 265, 22.8.1998.

OECD: Perspectives des communications, 1997.

OECD: Science, Technology and Industry Outlook, 1996, 1998.

Spectrum, Information Society initiative: 'Moving into the information society', an international benchmarking study, 1997.

# Annex

Table 9

## Main telecommunications directives and their implementation

Directive	Implemented in all countries at January 1998 except:
Framework 90/387/EEC as amended by 97/51/EC	(Belgium, Luxembourg)
Leased lines 92/44/EEC as amended by 97/51/EC	Greece, France, Italy, Portugal, Sweden (Belgium)
Voice telephony 95/62/EEC as revised by 98/10/EC	Greece, France, Ireland, Italy, Sweden (Belgium, Austria, Luxembourg, Netherlands, Portugal)
Licensing 97/13/EC	Greece, (Belgium, France, Italy, Luxembourg, Netherlands, Austria)
Interconnection 97/33/EC	Greece, Portugal (Belgium, France, Luxembourg, Sweden)
Terminals 91/263/EEC	
Satellite 93/97/EEC	Belgium, Greece, Ireland
GSM 87/372/EEC	Luxembourg
ERMES 90/544/EEC	Luxembourg, UK
DECT 91/287/EEC	Luxembourg

Source: Fourth report on the implementation of the telecommunications regulatory package, November 1998. Countries in brackets have implemented the Directives partially.

Table 10

**Network connection and supply times**

		Supply time target	Actual supply times
<b>Group 1</b>			
	Finland	1996: Within 5 days on average	1996: Within 3.8 days on average
	Sweden	Within 17 days	99 % within target
	Denmark	95 % on the date agreed upon with the customer. Remaining 5 % within 10 days.	96 % within target
	UK	BT has targets for: Residential: within 8 working days Business: within 6 working days	BT achieved: Residential: 76.3 % satisfied target Business: 70.3 % satisfied target
<b>Group 2</b>			
	Netherlands	Within 1 month	97 % within target, 91 % within 10 days
	Germany	80 % within 4 weeks (20 working days)	First half of 1996: within 10 working days: 77.4 % within 15 working days: 82.5 % within 20 working days: 85.5 % Second half of 1996: within 10 working days: 78.5 % within 15 working days: 81.4 % within 20 working days: 83.3 %
	France	Within 5 working days (by the end of 1998)	1995: within 7 days 1996: within 6.5 days
<b>Group 3</b>			
	Spain	Within 30 days	1995: within 3.3 days 1996: within 3.3 days
	Belgium	1995: 80 % within 5 working days 1996: 90 % within 5 working days	1995: 95 % within 5 working days 1996: 97 % within 5 working days
	Ireland	Within 11 calendar days	Within 11 calendar days
	Italy	For new connection: within 10 days <sup>(1)</sup>	1995: 98.7 % within target 1996: 97.8 % within target
	Austria	Within 7 days	Within 10 days
	Portugal	Within 1.6 month	Within 0.3 month
	Luxembourg	Within 20 working days (Target for 1998)	69 % within 30 days 98 % within 3 months 2 % over 3 months
	Greece	Within 1 week (in 2003)	In 92 % of exchanges less than 17 days

<sup>(1)</sup> New target established in 1997, previous target was within 60 days.

Source: Replies to Commission questionnaire for the first monitoring of universal service.

Table 11

**Fault repair: Targets and actual times**

Country	Fault repair target	Actual repair times
<b>Group 1</b>		
UK <sup>(1)</sup>	Business: BT has a target of 5 working hours. Residential: BT has a target of 9 working hours	Business: BT achieved 88.8 % within 5 working hours or by successful appointment. Residential: BT achieved 79.8 % within 9 working hours or by successful appointment.
Sweden	55 % within 8 working hours 100 % within 2 working days	75–80 % within first target 92–95 % within second target
Denmark	24 elapsed hours	25 elapsed hours <sup>(2)</sup>
Finland	1996: 78 % within working day	1996: 75.5 % within working day
<b>Group 2</b>		
Netherlands	Within 5 working days	99 % within target, 98 % within 2 working days
Germany	3 working days 2 working days (since 1/7/1995) 24 hours (since 1/7/1998)	First half of 1995: 85.53 % within target First half of 1996: 90.06 % within target Second half of 1996: 83.5 % within target
France	1995 and 96: 92 % the same day or before next working day in normal cases	1995: 88.3 %, 1996: 88.7 %
<b>Group 3</b>		
Spain	Within 15 hours	1995: 8.33 hours, 1996: 9.10 hours
Belgium	1995: 76 % before end of the next working day 1996: 80 % before end of the next working day	N/A
Ireland	18 hours	18.6 hours
Italy	Within 2 working days	1995: 67.4 % same day, 97.9 % within 2 days, 1996: 60.4 % same day, 95.4 % within 2 days.
Austria	Within one working day	93 % within 24 hours
Greece	95 % on the following working day (in 2003)	65 % the following working day
Luxembourg	Target for 1998: within 16 working hours	93.7 % same day, 100 % within 10 days
Portugal	85 % within 2 working days	80.5 % within 2 working days

<sup>(1)</sup> Supply times figures relate to the period October 1996 to March 1997; figures for the preceding six month are: residential 85.7 % and business 89.8 %. Other telecom companies, who are not universal service providers, report performance in 'Telecommunications Companies — Comparable performance indicators' published by P-E International

<sup>(2)</sup> The figure shows the number of finished fault repairs multiplied by the average repair time of all faults reported.

Source: Replies to Commission questionnaire for the first monitoring of universal service.

Table 12

Telecommunications penetration, 1996

	Lines per 100 inhabitants, 1997	% of households with a telephone	Share of digital main lines (%)	Mobile subscribers per 100 inhabitants, 1998
Sweden	68	98.7	96	44.5
Finland	56	95.0	95	50.4
Denmark	64	n.a.	84	28.4
UK	55	96.5	93	16.9
<b>Group 1</b>	<b>61</b>	<b>n.a.</b>	<b>90</b>	<b>35.1</b>
Netherlands	57	96.5	80	16.5
Germany	55	89.3	82	13.5
France	58	96.2	98	13.9
<b>Group 2</b>	<b>57</b>	<b>92.8</b>	<b>90</b>	<b>14.6</b>
Belgium	49	n.a.	(74)	13.1
Italy	45	87.9	85	28.7
Spain	40	87.5	67	13.9
Ireland	42	n.a.	83	18.5
Austria	46	n.a.	72	20.7
Greece	52	n.a.	43	14.1
Luxembourg	67	n.a.	(74)	20.8
Portugal	39	89.7	79	21.1
<b>Group 3</b>	<b>48</b>	<b>n.a.</b>	<b>65</b>	<b>18.9</b>

Source: EITO, p. 49 and European Commission: Fourth report on the implementation of the telecommunications package. November 1998. Figures in brackets cover Luxembourg and Belgium.

Table 13

GDP per head and telecommunications, 1996

	GDP per head 1996, ECU	Rank by GDP per head	Telecommunications expenditure, % GDP	Telecommunications expenditure per head, ECU
Sweden	22 534	4	2.69	597
Finland	19 314	9	2.46	469
Denmark	27 321	2	2.35	615
UK	15 476	12	2.84	429
<b>Group 1</b>	<b>21 161</b>	<b>6.8</b>	<b>2.60</b>	<b>528</b>
Netherlands	20 129	8	2.67	524
Germany	22 672	3	2.13	493
France	20 794	7	2.12	438
<b>Group 2</b>	<b>21 733</b>	<b>6</b>	<b>2.30</b>	<b>485</b>
Belgium	20 921	6	(2.27)	(470)
Italy	16 682	10	2.26	389
Spain	11 676	13	2.34	274
Austria	22 259	5	1.84	410
Ireland	15 750	11	3.78	540
Greece	9 228	14	3.05	269
Luxembourg	33 500	1	(2.27)	(470)
Portugal	8 253	15	3.29	283
<b>Group 3</b>	<b>17 284</b>	<b>9.4</b>	<b>2.70</b>	<b>376</b>

Source: Ameco, EITO. Figures in brackets cover both Luxembourg and Belgium.

Table 14

**Number of new jobs in 2005 resulting from liberalisation of telecom**

		Slow liberalisation (minimum liberalisation, restrictive national interpretation, remaining regulation)	Rapid liberalisation (maximum liberalisation, for instance telephony on CATV, capacity resale, alternative networks)
<b>Group 1</b>			
	Sweden	6 500	13 800
	Finland	7 900	8 100
	Denmark	13 900	14 700
	UK	97 900	178 000
<b>Group 2</b>			
	Netherlands	10 500	19 000
	Germany	59 200	153 000
	France	66 400	105 300
<b>Group 3</b>			
	Spain	41 800	55 900
	Austria	9 600	17 400
	Ireland	- 1 500	3 600
	Italy	40 900	33 000
	Belgium	2 800	1 500
	Portugal	8 500	11 500
	Luxembourg	800	800
	Greece	4 900	8 100

Source: BIPE Conseil: 'Effects on employment of the liberalisation of the telecommunications sector', January 1997. The results presented here relate to the scenario 'rapid technology diffusion' which assumes mass markets and high penetration rates.



# Part B

## Network industries and public service

Study carried out for the European Commission by the Institut d'Économie Industrielle, Université des Sciences Sociales de Toulouse





# Contents

<b>Introduction</b>	<b>67</b>
A. The economic characteristics of network industries	67
B. Rationale for and implementation of industry oversight	70
C. Universal service obligations	76
D. Policy implications for European network industries	78
<b>Part one: Deregulation of network industries and the role of the public authorities</b>	<b>81</b>
<b>I. Network industries and network goods</b>	<b>81</b>
Introduction	81
1. The concept of network	81
1.1. Definitions	81
<i>Box 1: Dense versus minimal network</i>	82
1.2. Some illustrations	83
1.3. Operators and users	84
2. Costs and benefits from integration	85
2.1. Economies of scale and economies of scope	85
<i>Box 2: The gains from large scale operation</i>	86
2.2. Vertical integration	87
3. Density, externalities and interconnection	87
3.1. The economies of density	88
<i>Box 3: The gains from rerouting</i>	90
3.2. Clubs and networks	91
<i>Box 4: Network externalities and multiple equilibria</i>	92
4. Networks and competition	93
4.1. Competition under network externalities	93
4.2. Investment and dynamic competition	95
5. A typology of network industries	95
	59

<b>II. Government intervention in network industries</b>	<b>99</b>
Introduction	99
1. Modes of government intervention	99
1.1. Legislate and enforce the law	99
1.2. Tax and subsidise	100
1.3. Sell, rent and buy	100
1.3.1. The problem	100
1.3.1.1. Examples	100
1.3.1.2. Problems in the contracts	100
(1) Choosing a firm	101
(a) Unknown costs	101
(b) Unknown demand	101
(c) Unknown willingness to pay	101
(d) Collusion	101
(2) Contractual difficulties	101
(a) Measuring the performance of the firm	101
(b) Allocation of risk	102
(c) Investment and maintenance	102
(d) Modifying the terms of the contract	102
(e) Renegotiation and the ‘hold up’ problem	103
(f) The incentives of the government	103
1.3.2. Techniques	103
1.3.2.1. What type of contracts?	103
(1) Introducing competition	103
(2) Contracting for specific services	103
(3) Regulation	103
(4) Franchising	104
1.3.2.2. Choosing the firm(s)	104
(1) Negotiations	104
(2) Auctions	104
1.4. Administer	105
2. Correcting market failures	105
2.1. Efficiency and markets	105
2.2. Promoting competition	105
2.2.1. Antitrust	105
2.2.2. Promoting entry	106
2.3. Allocating property rights	106
2.3.1. The problem	106
2.3.2. Methods for allocating property rights	106
2.4. Managing natural monopolies	106
2.4.1. Technology	106
2.4.1.1. Traditional theory	106
2.4.1.2. Scope of the argument	106
(1) Competition in some segments and not in others	106
(2) Managing the interface	107
(3) Contestable markets	107
2.4.2. Contracts	107
2.4.2.1. Incomplete contracts and vertical integration	107
2.4.2.2. Managing the contracts to promote competition	108

2.5. Information management	108
2.6. Failure of the capital markets	108
2.7. Network externalities	108
2.8. International considerations	109
2.8.1. Political risk	109
2.8.2. Using market power	109
3. Income distribution	110
3.1. Markets and income distribution	110
3.1.1. The problem	110
3.1.2. Taxation	110
3.1.3. Free provision of goods	110
3.1.4. Application to network industries	110
3.2. Merit goods	110
3.3. Regional policy: income distribution among regions	111
4. The costs of government intervention	111
4.1. The efficiency of public institutions	112
4.2. Capture	112
4.3. Income distribution from the poor to the rich	112
<b>Appendix: Public versus private firms — a survey of empirical tests of relative performance</b>	<b>113</b>
1. Delineation	113
2. Defining and measuring economic performance	114
3. The production frontier approach	116
4. Survey of empirical studies	118
5. Conclusion	119
Note	119
<b>III. Competition policy versus regulation</b>	<b>123</b>
Introduction	123
1. The objectives of industry oversight	123
2. Procedures and control rights	124
3. Timing of oversight	125
4. Information intensiveness and continued relationship	126
5. Independence vis-à-vis the political environment	127
6. Toward the demise of regulation?	128
7. Are antitrust and regulation substitutes or complements?	130

<b>IV. Pricing interconnection</b>	<b>132</b>
Introduction	132
1. Pricing interconnection of competitive services to a monopoly infrastructure network	133
Introduction	133
— Background	133
— Interconnection policies	134
— Conceptual framework	135
— Workability of the optimal access pricing rule	135
1.1. A theoretical framework	135
1.2. Fully-distributed costs	136
1.3. The OFTEL rule	137
1.4. The efficient component pricing rule	137
1.5. The optimal access price	139
1.6. Is ECPR consistent with an optimal access pricing rule?	139
1.7. A possible implementation of usage-based access prices: the global price cap policy	140
1.7.1. Partial versus global price caps	140
1.7.2. Global price cap with and without ECPR	141
1.8. Lack of instruments and multiple goals for interconnection charges	142
1.8.1. Competitors' market power	143
1.8.1.1. Profitable entry	143
1.8.1.2. Unprofitable entry	143
1.8.2. Bypass and network duplication	143
1.8.2.1. Bypass	144
1.8.2.2. Network duplication	144
1.9. Access charges and incentives for cost minimisation	145
1.9.1. Incentives to minimise bottleneck cost	145
1.9.2. Incentives to minimise cost on the competitive segment	145
Conclusion	145
2. Bottleneck access and market foreclosure in an unregulated environment	147
Introduction	147
2.1. What is foreclosure?	147
2.2. Envisioned remedies	148
2.3. The theory of market foreclosure	150
2.4. Policy and business strategies implications	151
2.5. Potential defences for exclusionary behaviours	151
3. Two examples: transportation in the gas and the electricity industries	153
4. Competition between infrastructure operators and the problem of two-way access	156
Introduction	156
4.1. Network competition in a mature industry	157
4.1.1. Uniform price competition	158
4.1.2. Competition in two-part tariffs	158
4.1.3. Price discrimination based on call termination	159
4.1.4. The determination of access charges	160
4.2. Entry in network competition	161
References	163

<b>Part two: Universal service</b>	<b>168</b>
<b>I. Universal service: definition, cost and financing</b>	<b>168</b>
1. Introduction: Public service versus universal service	168
1.1. Public service	168
1.2. Universal service	170
2. Fundamentals	171
2.1. Definition and economic content	171
2.2. Rationale	172
2.2.1. Network externalities	172
2.2.2. Redistribution	173
2.2.3. Public/merit good	175
2.2.4. Others	175
2.2.4.1. Regional policy	175
2.2.4.2. Partisan politics and pressure groups	176
3. Cost of universal service	176
3.1. Definition	176
3.1.1. Profitability cost	176
3.1.2. Welfare cost	177
3.2. A broader view: evaluating costs and benefits of universal service	177
4. Implementation and financing	178
4.1. Monopolistic sector	179
4.2. Liberalised sector	180
4.2.1. The universal service obligation is imposed on a single, specified operator	181
4.2.1.1. The operator who faces the USO is solely responsible for the financing	181
4.2.1.2. All operators contribute	181
4.2.2. Franchising of the universal service obligation	183
<b>II. Universal service in practice — some experiences</b>	<b>185</b>
1. Introduction	185
2. The definition of universal service	185
2.1. Telecommunications	185
USA	185
UK	187
Germany	187
2.2. Postal service	188
EU	188
2.3. Railway transport	189
EU	189
2.4. Water	189
EU	189
USA and Japan	190
2.5. Electricity	190
2.6. Gas	191

3. The financing of the universal service	191
3.1. Telecommunications	191
EU	191
UK	191
3.2. Postal service	195
EU	195
Germany	195
Sweden	196
3.3. Water	196
EU, USA and Japan	196
References	197
<b>Part three: Main policy issues in network industries</b>	<b>198</b>
Telecommunications	198
1. Key technological specificities	198
2. Policy issues in the transition to full network competition	198
3. Long-run policy issues	199
4. Current issues facing the Commission	199
Postal service	200
1. Technological specificities	200
2. Policy issues	201
Electricity	202
1. Technological specificities	202
2. Policy issues	203
Natural gas	204
1. Technological specificities	204
2. Policy issues	205
Water	206
1. Key technological and economic specificities	206
1.1. Local sub-networks and scale economies	206
1.2. Global management of the water resource	206
2. Policy issues	207
2.1. Regulation of local networks	207
2.2. Global management of the resource	207
Urban transport	207
1. Characteristics	207
1.1. Technology	207
1.2. Regulation	208
2. Policy issues	208
2.1. Liberalisation and regulation environment	208
2.2. Environmental issues, in urban transportation technologies, and intermodal competition	209
Air transport	209
1. Technology and networks	209
1.1. Airlines	209

1.2. Airports	210
1.3. Air-traffic control	210
2. Policy issues	210
2.1. Airlines	210
2.2. Air-traffic control	211
2.3. Airports	211
Railway transport	
1. Specificities	211
2. Policy issues	213
References	215





# Introduction

The main purpose of this report is to provide a conceptual framework in which economic issues related to the deregulation of network industries can be analysed dispassionately. It is composed of three parts. After a discussion of the rationale for the oversight of network industries and its alternative modes (competition policy versus regulation), Part one focuses on the design and implementation of interconnection policies. Part two analyses alternative ways of providing universal service. Last, Part three outlines the basic technological features that characterise each industry and draws attention to the main economic policy issues.

## **A. The economic characteristics of network industries**

### **I. Network industries: a wide array of situations ...**

The network industries reviewed in this report differ in many respects. The first obvious difference corresponds to their GDP share: Some industries are quite important (for example, the telecommunications industry represents approximately 2 % of GDP) while others are much smaller (for example, the sector of urban transport represents less than 0.4 % in most countries). They also differ in their intensities of labour and capital investment (the telecommunications industry represents only 0.7 % of total employment whereas for postal services labour costs account for 80 % of the operators' costs), as well as in their profitability: the 25 top European telecom operators are more profitable than the top 100 banks, whereas throughout Europe railways companies face financial difficulties, and urban transport services are heavily subsidised (between 25 % and 75 % of their operating costs).

Next, these industries differ in the rate of evolution of their technologies. The telecommunications industry provides again an extreme example. Extremely rapid technological progress has increased its productivity

and led to a decrease in real prices, due in part to sharp decreases in the cost of transmission and switches. It has also affected the structure of the industry, for example with the appearance of some competition for local services from operators using cable and wireless technologies. To a more limited extent, the introduction of new technologies in electricity generation (gas generating units, cogeneration) has also affected the structure of the industry by reducing the efficient size of generating units. In contrast, the evolution of other sectors is much slower and some of them seem to be driven mostly by the technological evolution of competing services. For example, although there has been some significant amount of technological progress in the provision of postal services (introduction of optical character readers, remote barcode sorting, and delivery point sequencing), the potential for future technological innovations appears to be limited; and although the volume of communications has doubled between 1974 and 1994, the expansion of wire communications (fax, e-mail) has reduced the share of postal administrations in the market for communications. As a result, whereas the demand for telecommunications is increasing, the demand for postal services is expected to be at best stable, and probably declining. Likewise, in spite of the introduction of high-speed trains, the evolution of the rail transport industry is primarily driven by increased competition from alternative means (road and air) of transportation.

Lastly, the adoption of technological standards is not as advanced in all industries. For example, whereas the adoption of the GSM technology ensures that the same handset can be used throughout Europe for mobile telecommunications services, technical compatibility problems severely hinder the interoperability of the European railway networks: there are four different rail gauges, different gabarits for the rolling stock (e.g. bridges and tunnels are smaller in the United Kingdom than in continental Europe), six different types of electric power with no locomotive being compatible with

more than two types; and the transmission systems between tracks and locomotives are country-specific.

## **II. ... and many common characteristics**

Despite those important differences, the network industries under review share common characteristics. First, they all involve several activities, some of which have the characteristics of a natural monopoly and are typically subject to some form of regulation, some of which may potentially be more open to competition.

The telecommunications industry may again constitute an extreme example, since some services have already been open to competition for some time now, for example long-distance services in the USA and the UK, mobile communications in the USA and in several European countries, local services in the UK and, somewhat more recently, in Finland, Japan, Sweden and Denmark. The entire industry will be fully liberalised by 1 January 1998. (However, it should be noted that although there more than 400 active firms in the USA long-distance and international sectors, these two sectors are still dominated by the former monopoly, AT&T. Likewise, in the other countries and/or segments, the former monopoly still often dominates a relatively concentrated market.) Note that competition may involve duplication of the infrastructure (fibre optic cable for long-distance telecommunications in the USA, in England and Japan, and probably in continental Europe after 1998; cable and wireless operators for local services, although the local loop is still perceived as a bottleneck) or require leasing the entrants of parts of the incumbent's infrastructure (resale or unbundling). In all cases, competing networks need to be interconnected for customers to benefit from 'network externalities'.

The water industry may illustrate the opposite polar case in which large-scale competition in the market is unlikely to develop in the near future: since transportation costs are relatively high, there exists a large number of local networks (cities, metropolitan areas, counties, etc.) which are usually not interconnected, except in cases where it is necessary to provide water to regions with a particular need, as for example for the Flanders region in Belgium and for some Spanish provinces. Each local network is in fact composed of two sub-networks: a distribution network that provides water to users, and a water purification network that collects used water, purifies it, and recycles it through the network or discharges it into watercourses. Natural

monopoly seems to be the fundamental nature of each sub-network, and there does not seem to exist scope economies between those two sub-networks. However, pumping, billing and revenue collection might be organised in a competitive way. Yet competition for the market (in the form of an auction) is likely to remain the dominant mode of competition in this industry.

In postal services, four different activities can be identified: collecting, sorting, transportation and distribution; while the distribution stage has most of the characteristics of a natural monopoly, there is already some competition in express services with high value added and, in the future, more competition may appear for parcel services but also in the booming industry of business mail for private individuals (both addressed and non-addressed advertisement — a low value-added segment).

In the electricity industry, due in part to new technologies, which have reduced the minimum efficient scale of generating plants, generation is generally viewed as being potentially competitive whereas the transmission and the distribution sectors have the characteristics of natural monopolies. A similar distinction applies to the gas sector, although the number of producers is very limited (mainly Russia, Norway, Algeria and the Netherlands); there may be potential competition in the provision of seasonal storage facilities, such as the depleted gas field in the North Sea, although geographical and historical conditions may make existing storage an essential facility, to which suppliers must have access in order to compete effectively. In contrast, storage facilities intended to smooth daily peaks are an integral part of the transportation system and cannot be easily unbundled from transportation services.

In air transport, two different markets might be distinguished, one for passengers and one for freight, with the former generating most of the industry income. Whether there are significant returns to scale depends on the types of routes: for long haul flights, operating two planes from one airport to another may not cost much less than twice the cost of operating only one plane between those two airports. However, the marginal cost of a passenger on a flight is close to zero, and hence the cost function has jumps. This explains the importance of the drainage of consumers to long distance flights and the development of feeder routes. On very small routes, the situation is somewhat different as

the fixed costs of a minimum crew and ticketing facilities already represent a sizeable portion of the total expected revenues, and increasing returns to scale thus become important. Also, consumers are sensitive to the frequency with which an airline flies between two airports (hence the development of shuttle services, such as the new SAS Express service), and to the possibility of easy connection (an airline with many routes thus has an important competitive advantage — hence the introduction of codesharing). Airport capacity is another constraint in air transportation; physical and environmental limitations on the supply of infrastructure also place limitations on the supply of operations, thus creating a further need for special regulation to ensure ‘fair and equal opportunity’ of access to this limited supply.

For rail transport, passenger and freight services must again be distinguished. In 1991, a first directive <sup>(1)</sup> imposed accounting or organic separation between transport operations, which are potentially open to competition, and infrastructure activities which are the natural monopoly segment. In 1995, a second directive <sup>(2)</sup> further required the infrastructure to be monitored by an independent body to ensure non-discriminatory access to the infrastructure.

Two series of issues derive from these characteristics. How shall one design proper incentives to maintain, operate and invest in the natural monopoly segments, and how should competition in the potentially competitive segments be encouraged?

Regarding the first type of issue, experience suggests that both the ownership structure and the type of regulation matter. Consider for example the case of urban transport, where the situation varies much from one country to another, and where moreover substantial changes have occurred in some countries in the last two decades. In France, operators have generally higher costs under a cost-plus regime regardless of their ownership structure (private, semi-public or public). In the UK, the 1985 British Transport Act deregulated and privatised urban transport; large cost savings have been achieved and subsidies have been reduced, but fares have risen while passenger trips and revenues have decreased (the experiment seems to have been more successful in London than in the other metropolitan areas, and for buses than for urban trains). In Japan, full

deregulation of urban transport, and in particular of urban rails, appears successful to most analysts.

The analysis of the monopoly segments raises several questions. Should the operators present in the various segments divest themselves and, if yes, should there be lines-of-business restrictions? For example, should the operator in charge of the high-voltage grid be allowed to generate electricity or to distribute it? Should the operator in charge of maintaining the railway network be allowed to offer rail transportation services? The basic trade-off here is between efficiency gains (from enhanced coordination between the investments in generation and in the transmission grid, say, assuming that this coordination is better achieved through vertical integration) and the necessity to ensure a level playing field for the operators active in the competitive segments.

The transition from a pure monopoly position to a more competitive situation also poses specific problems, particularly when the industry is labour intensive and incumbent operators are overstaffed.

A second characteristic that is common to many network industries is the existence of universal service obligations (USOs). These may take various forms: provision of a minimal service or of a minimal quality at a reasonable or affordable price, tariff per equation, for example a uniform price across regions or across consumer groups, in spite of differences in the cost of providing the service, etc. For example, for postal services the Directive 97/67/EC <sup>(3)</sup> specifies that USO should include the clearance, transport, sorting and distribution of postal items up to 2 kg and of postal packages up to 10 kg, as well as services for registered and insured items. Universal service obligations may include social tariffs or subsidised services for some groups of consumers; they may apply to all consumers or only to existing ones (excluding new customers), etc. The cost of these obligations may also be financed in a variety of ways, e.g. through internal cross-subsidies between an incumbent’s product prices, through access charges, through direct subsidies from a universal service fund, or by giving the operator a monopoly situation in some segments. For postal services, for example, USOs are financed through cross-subsidies combined with the exclusive reservation of some markets for services that

<sup>(1)</sup> Directive 91/440/EEC of 29 July 1991 (OJ L 237, 24.8.1991).

<sup>(2)</sup> Directive 95/19/EC of 29 June 1995 (OJ L 143, 27.6.1995).

<sup>(3)</sup> OJ L 15, 21.1.1998, pp. 14–25.

are included in the universal services basket. The appropriate definition of the USO and the less distorting way to finance it are the two key issues in this matter.

Some characteristics are shared by only a subset of network industries. For example, both the electricity and the gas industries face a demand management problem, as the demand exhibits both a seasonal pattern, which may not be easily matched by the producers (even though they can use seasonal storage facilities as in the case of gas), and an aggregate uncertainty (due in particular to the uncertain weather), which may require some coordination between the suppliers' investments in extra capacity and the customers' investments in substitute energy sources. Other industries still face underdeveloped networks — this is for example the case of Greece, Spain, Portugal and the eastern part of Germany for telecommunications infrastructures, and of several Mediterranean countries for water purification infrastructures. Most of the transportation industries face congestion problems, which raise efficiency issues, since users impose costs on others by increasing their waiting and transfer time, but also intermodality and environmental issues. Lastly, many industries generate public good problems, for instance the availability and the quality of water, which raise difficult coordination problems since different regions or countries obtain their water from common sources, or the safety of air transportation.

## **B. Rationale for and implementation of industry oversight**

### **I. The concept of network industry**

While engineers tend to define networks as a set of nodes and interconnecting lines between the nodes, economists prefer to focus on those features that might interfere with an efficient allocation of resources by an unregulated private sector. The various industries analysed in the report are too disparate to be encompassed in a single paradigm. Yet most of them possess a couple of features that have in the past motivated the introduction of regulation in lieu of the standard antitrust oversight. They typically involve segments that exhibit substantial returns to scale and therefore are technologically natural monopolies. Returns to scale may involve an infrastructure whose unit cost sharply decreases with volume or whose operation requires an integrated management: an electricity transmission grid, mail collection, or a water or sewerage network. Technological considerations suggest that the infra-

structure ought not be duplicated. This in turn implies that customers and service providers must be given access to the infrastructure. Sometimes, returns to scale may also occur at the customer level as it may be inefficient to connect the customer to several companies. Such scale economies at the customer level may also create a need for interconnection if simultaneously there are network externalities among customers, that is, if a customer enjoys the service more when other customers also consume the same service. This brings us to a second feature present in some of the industries covered by this report. Consumers may want to combine services supplied by several operators. A subscriber of a telecommunications operator generally wants to call subscribers of other operators and therefore use off-net termination services. A passenger of an airline, train or bus company may want to take a connecting flight, train or bus provided by a different company. A letter or a package may be shipped through several complementary postal services.

The existence of infrastructures and of network externalities raises the issue of compatibility and interconnection among non-affiliated operators. Accordingly, the report devotes substantial attention to the interconnection issue.

### **II. Government intervention in network industries**

Four roles have traditionally been assigned to the State with respect to its oversight of industries. First, the State should provide a legal framework that enables an efficient and reliable enforcement of private contracts. It should also define general rules of conduct that are applicable to most or all industries. Antitrust statutes are meant to put broad restrictions on the formation of cartels, on the abuse of dominant position, and so forth. Second, the State should substitute for missing private contracts; the typical example is the case of diffuse pollution, where the number of polluters makes it prohibitively costly to design private contracts that force potential polluters to internalise the full social cost of their actions; in the same vein, government intervention (in the form of taxes, quotas or markets for pollution permits) can prevent rent seeking by potential polluters who could install facilities and then search for monetary compensation for not polluting. Third, the existence of infrastructures and/or network externalities raises a concern about the efficient functioning of markets, even if those are freed from collusive or predatory practices by standard antitrust enforcement. A third role of State intervention is thus to promote productive efficiency

without creating excessive rents and to offer a satisfactory array of services at reasonable prices to consumers. Fourth, the State may want to depart from pure economic efficiency for redistribution purposes. The State may then redistribute income or services across consumers or across geographical areas.

The report's view is that the first two aspects of State intervention into industries are fairly orthogonal to the nature of network industries. In contrast, the last two roles, while not specific to network industries, are prominent in such industries, and therefore are the focus of the rest of the report

### III. Competition policy versus regulation

The opening of network industries to competition has paved the road for a possible replacement of the traditional regulatory paradigm by antitrust policy. In the previous monopolised environments, the main economic issues were the setting of a proper price structure, the encouragement of cost efficiency, the extraction of monopoly rents and the protection of the company's long-term investments against expropriation. Antitrust policy, with its focus on competition and conduct, had little to contribute. The prospect that the emerging competition in network industries may discipline firms and force them to be efficient, may rebalance price structures, and may limit potential rents, and the fact that competitors will constantly interact at the wholesale and retail levels clearly raise the institutional design questions of which of regulation and competition policy is best placed to provide the oversight of network industries, and of whether these two modes are complements or substitutes.

In an attempt to clarify the debate on who should oversee the liberalised network industries, the report begins with the three standard hazards faced by government intervention into industries: imperfect knowledge of cost and demand conditions, capture by interest groups and political influence, and limited commitment ability. After stressing the points of convergence between regulation and antitrust, it emphasises several distinctions between their standard modes of operation. It is argued that the choice between the two institutions affects: (a) the transparency of redistributive policies (as less visible forms of cross-subsidies are less likely to be performed in a deregulated framework); (b) the timing of oversight (antitrust enforcement usually operates *ex post* while regulation allows for more *ex ante* decisions; an *ex ante* treatment enhances commitment, removes part of the

uncertainty faced by firms for their investment decisions, and is more expedient; but it may result in policies that do not reflect future information acquired by government officials, and it may further encourage capture); (c) the nature of the information on which the oversight is based (regulators are often better informed than antitrust authorities; this however has a cost, as, first, superior information may worsen the time consistency problem of public policies, and, second, regulatory expertise is usually acquired through a long-term relationship with the industry, which may facilitate capture); and (d) the role and incentives of interveners (that is, the benefits that the interveners may expect when bringing information that may impact on decisions).

The report also discusses the issue of independence of the industry overseer with respect to the political system. It reminds the reader of the standard trade-off between accountability, which is obtained by letting elected officials control the overseer, and integrity, which is jeopardised by the capture of elected officials by interest groups. To the extent that decisions that have to be reached in the context of network industries are usually technical and poorly understood by the electorate, the accountability argument should be given less weight than in other contexts, and so the argument for independence is reinforced.

The report suggests that traditional stand-alone antitrust enforcement may in most network industries not provide the smooth competitive environment it is supposed to create. As already mentioned, a key feature of a number of network industries is the need for mutual interconnection among competitors. While it is straightforward to mandate open access on paper, it is much harder to confront the subtle issues involved in developing compatibility and interconnection: technological requirements, level of unbundling, quality and timing of interconnection and level of interconnection charges. Each of these dimensions can give rise to anti-competitive foreclosure behaviours by incumbents on the one hand, and to excessively costly demands by entrants on the other hand. Designing a proper interconnection policy requires not only: (a) a sophisticated understanding of economic incentives and effects, but also, and to a varying degree depending on the regulatory mode, (b) substantial technological expertise, and (c) considerable cost and demand information. Although courts occasionally investigate cases of alleged foreclosure by an essential facility owner in unregulated environments, they are unlikely to engage in detailed oversight of



interconnection policies. And indeed, as illustrated by the Clear-Telecom dispute in New Zealand, courts have been somewhat reluctant to become involved in such policies. The *ex post* nature of competition policy also creates considerable uncertainty for entrants who have to sink substantial investments without knowing an important determinant (the interconnection charge) of their profitability. This further handicap of competition policy in the matter of network industry oversight is even heavier when the entrants need to leverage themselves up in order to finance the said investments, since the capital market may be concerned by the enhanced riskiness of lending. Of course the report does not argue that antitrust enforcers are unable to handle foreclosure issues; and indeed they do so in a number of unregulated industries, as noted in the discussion of vertical integration, tie-ins, refusal to deal or to cooperate, and price discrimination later in the report. Rather, the report emphasises the need for a careful analysis of whether standard antitrust intervention against foreclosure or tacit collusion among competitors through voluntary wholesale agreements is appropriate in each of the industries covered by this report.

Last, the report discusses whether regulation and antitrust are substitutes or complements. The case for substitution (and thus for a unified oversight) rests on the possibility of inconsistent or uncoordinated intervention by multiple overseers. But a strong case can be made in favour of the coexistence of the two modes of oversight. This case can be based on a reduced scope for capture, on a smaller likelihood of cover ups in case of regulatory mistakes, and on the usefulness of creating advocates for specific causes (for example, the antitrust enforcers being advocates for competition).

#### **IV. Pricing interconnection**

In many network industries, the key to the creation of competition is the definition of an interconnection policy. The report considers three prominent situations. The first two (Sections 1 and 2 below) look at one-way access to an essential facility in a regulated and an antitrust context, respectively. The third one (Section 3) corresponds to a two-way access between two bottleneck owners whose services must be interconnected.

##### *1. Pricing interconnection of competitive services to a monopoly infrastructure network*

A major policy issue in network industries is the liberalisation of potentially competitive segments which need

the network as an essential input, which is thus an 'essential facility' or 'bottleneck'. The essential facility is often monopolised because of large economies of scale, of first-mover advantages or of technological superiority. The policy-maker must induce an efficient allocation of resources. This involves, among other things, creating proper conditions for entry into the competitive segment while not inducing excessive entry, not expropriating previous investments or discouraging future ones in the monopolised segment, or not generating inefficient bypass.

This question is a classic one in antitrust economics, as courts have been asked to investigate the existence of foreclosure and the design of relief policies in industries as diverse as the cement, railroad and computer reservation systems industries. This question has also lately received substantial attention in the regulatory context of the creation of competition in various segments of the activity of a dominant natural monopoly. In the telecommunications industry, the impetus for the development of interconnection policies was the opening of competition in the long distance markets; new entrants (such as Mercury in the UK and Sprint and MCI in the USA) needed access to the dominant operators' (British Telecom, AT&T) local networks to reach the customers; it then became clear that interconnection charges would play a much broader role as competition started developing in the local segment from cable companies and mobile operators who needed access to long distance services and in value-added services, and as networks began to proliferate. But the issue is not specific to telecommunications: it has been argued for instance that transmission for electricity, pipelines for gas, tracks, rails and stations for railroads, and mail collection and distribution for post offices are natural monopoly segments to which rivals in complementary segments (generation of electricity or production of gas, freight or passenger train operation, mail transportation) must have access in order to compete.

The recent years have witnessed a broad array of recommendations and practices. Entrants typically argue for a cost-based access charge such as long run incremental cost of access. It is however, widely accepted that marginal cost pricing of access prevents the dominant telephone operator from efficiently recovering the fixed costs of the network (and possibly the deficit stemming from the universal service constraints). Many regulators and economists have suggested long-run incremental cost plus a mark-up as a workable alterna-

tive to allow recovery of the 'access deficit'. What is put into the mark-up is the object of intense debate. Should it be an accounting allocation of the access deficit as under fully distributed costs? If so, should it take the form of a uniform mark-up on all services or else of a mark-up proportional to long-run incremental cost (the 'Allais rule'), as was suggested in the WIK-EAC report written for the European Commission or recommended, in the form of Forward Looking long-run incremental costs (FLLRIC), by the 1996 US Telecom Act? Or should the mark-up be related to the use that is made of the access by the telephone operator's competitors and therefore depend also on demand considerations?

The report first reviews standard interconnection policies and states the conditions for efficient one-way interconnection. It concludes that:

- (a) A discussion of an access rule without reference to the rest of the regulatory environment has limited interest. The quality of an access pricing rule depends on the determination of prices for the final products.
- (b) If the regulators are not constrained in their panoply of instruments, the access charge is determined solely by the concerns of creating efficient entry and avoiding inefficient cream-skimming, and of contributing to the recovery of the access deficit. The 'generalised access charge' (access charge, plus excise tax on the competitors' goods) depends on the use that is made of the access.
- (c) The dominant, practically-oriented view is to base the pricing of the infrastructure owner's intermediate goods purely on costs. This approach includes the traditional fully-distributed-cost (FDC) methodology. Basing access prices on cost may have perverse incentive implications. Further, fully distributed cost pricing induces cherry-picking by the competitors, and thereby generates allocative inefficiencies and jeopardises the recovery of the access deficit. It also creates inefficient bypass.

The Ramsey approach, which limits incentives for cherry-picking, is theoretically more satisfactory. The main difficulty with this approach is that the regulators are unlikely to hold the required informa-

tion about demand. The main hope for obtaining Ramsey-orientated prices is the delegation of pricing to the operator. A partial price cap, that is a price cap on retail services only, has two flaws compared to a global one, encompassing both retail and wholesale services. First, the omission of the intermediate goods in the operator's price cap subsidises price increases on the competitive segments and biases the rate structure toward high prices in those segments and low prices in the non-competitive ones. Second, this omission complicates the setting of the weight of competitive prices in the price cap, as it requires forecasts of the operator's market share besides the prediction of total demand in the competitive segment. A global price cap penalises increases in both access prices and final prices and induces the operator to price discriminate very much the way an unregulated firm would do, except that its entire price structure is brought down by the cap.

- (d) The separation of powers has deprived regulators of many useful instruments such as excise taxes on competitors' products, profit redistribution, or lump-sum entry subsidies. Consequently, regulators are forced to use the access price to arbitrate among conflicting goals (recover the access deficit, prevent inefficient bypass and network duplication, capture competitors' rents or ensure that competition is viable, etc.). The report describes the direction of the corrections that must be made in the access rule to reflect the lack of instruments.

## 2. *Bottleneck access and market foreclosure in an unregulated environment*

Where an unregulated monopoly operates a facility that is an essential input to competitive service providers, the situation raises market foreclosure issues, which has given rise to the so-called essential facility doctrine. (Examples of essential facilities to which competition law has been applied include a stadium, a railroad bridge or station, a harbour, a power transmission or a local telecommunications network, and a computer reservation system.) According to the received definition, foreclosure refers to any dominant firm's practice that denies proper access to an essential input it produces to some users of this input, with the intent of extending monopoly power from one segment of the market (the bottleneck segment) to the other (the potentially competitive segment). Strongly criticised by the



Chicago School, the foreclosure doctrine has recently been revived by the observation that such an upstream monopolist in general cannot fully exert its monopoly power without engaging in exclusionary practices: it thus has an incentive to exclude some downstream users, not to extend its monopoly power in the downstream segment, but to preserve and exert its own upstream monopoly power. (A similar observation is commonly made for patent licensing and franchising. A patent-holder, for example, is unlikely to make much money if it cannot commit not to flood the market with licences: if everyone holds a licence, intense downstream competition destroys the profit created by the upstream monopoly position; therefore, a patent-holder would like to restrict the number of licences. A similar argument applies to the number of franchisees.)

Foreclosure can be complete (as in the case of a refusal to deal) or partial (as when the bottleneck owner simply favours some downstream firms, e.g. its subsidiary), and can also be performed in various ways: (a) the bottleneck owner can integrate vertically with one or several firms in the complementary segment and refuse to deal with potential competitors (relatedly, it may engage in tie-ins and refuse to unbundle, thereby denying access to the essential facility); (b) in the presence of economies of scope or scale generated by the cooperation between firms in the same market, a dominant group of firms may put its competitors at a disadvantage by refusing to cooperate; (c) the bottleneck owner can grant exclusivity to a subset of firms on the complementary segment, and thus de facto exclude their rivals; (d) second- and third-degree price discrimination generalises exclusivity arrangements by favouring some customers over the others, while giving the bottleneck owner some flexibility in serving discriminated-against customers.

A number of remedies have been considered, which can be grouped into five categories: (a) structural policies such as divestitures and line of business restrictions are often considered in last resort, as they may involve substantial transaction costs of disentangling activities and may jeopardise the benefits of integration; (b) access price or quantity control; (c) price linkages: antitrust authorities often try to use other prices — for access or retail goods — as benchmarks for the access price (the efficient component pricing rule (ECPR), also called the Baumol–Willig rule, the imputation rule, the parity principle, and (perhaps confusingly) the non-discrimination rule, links the integrated monopolist's access and

retail price: namely, the access price charged to competitors should not exceed the price charged by the integrated firm on the competitive segment minus the incremental cost of that firm on the competitive segment); (d) 'common carrier' policies; and (e) disclosure requirements.

Three broad policy implications can be drawn from the recent literature. First, it does matter whether the more competitive of two complementary segments lies 'upstream' or 'downstream': prices are always lower when the bottleneck owner lies upstream (that is, is not at the interface with final users: this observation thus supports common carrier policies). Second, non-discrimination laws may have the perverse effect of restoring the monopoly power that they are supposed to fight (non-discrimination rules benefit the upstream bottleneck because, by forcing it to sell further units at the same high price as the initial ones, they help the bottleneck commit not to flood the market). Third, the efficient component pricing rule, which was designed for a regulated environment but is also used in antitrust contexts, often has little bite in an unregulated environment: ECPR only provides a link between access and final prices and is therefore only a partial rule; an integrated firm with upstream market power can still exercise its market power by setting a high price for the final good and, at the same time, set a high access charge to prevent other firms in the competitive segment from becoming effective competitors.

Vertical integration and foreclosure may also have social merit in some instances. For example, unrestrained competition may sometimes lead to excessive entry and duplication of fixed costs, and vertical foreclosure may help reducing this excessive entry. Also, vertical integration may help the upstream and downstream firms to achieve a better coordination, for example by providing better incentives to monitor firms' efforts. Last, foreclosure, like patents, may provide an innovation with a rent and therefore an incentive to innovate. The dynamic efficiency gains linked with innovation must then be traded off against the static loss created by foreclosure. This part of the report concludes with a list of such potential defences.

### *3. Competition between infrastructure operators and the problem of two-way access*

In many network industries a developing liberalisation has already produced, or is likely to produce, substantial changes in the way a number of the industries cov-

ered by this report operate. In particular, in some of those industries, regulatory scrutiny has already given way (or soon will) to a competitive market place from which detailed regulation has withdrawn. This is for example the case for air transportation, which was liberalised long ago in the USA and will soon be so in Europe. Similarly, in telecommunications, legislators, regulators, and antitrust authorities envision a transitional period followed by the substitution of competition policy for regulatory supervision: in the USA the 1996 Telecommunications Act aims at facilitating entry (including by the long distance companies) into the regional Bell Companies' territories while allowing the latter to enter the long distance market; in Europe the industry will be fully open to competition by January 1998. New Zealand's 'light-handed' regulatory regime, which relies on private negotiations between competitors to secure interconnection agreements, already provides a test laboratory: the regulatory authority has been abolished and the dominant firm (Telecom) and the entrants (Clear, BellSouth New Zealand) are subject only to general antitrust provisions on the abuse of dominant position and on vertical restraints.

When several infrastructure operators compete head-to-head for final customers, a first issue, already mentioned in Section 1 above, concerns the compatibility of the competing networks. There exists a second and related issue: even when network operators choose or are required to be compatible or interconnected, there remains the issue of the determination of the mutual access conditions provided by the competing operators. Two main concerns arise in this context.

First, one may fear that established networks use their interconnection agreements to facilitate cooperation in the final market (collusive behaviour concern). This concern is validated by recent analyses of competition between established telecommunication networks, which compete for subscribers but, at the same time, will have to agree on mutual access conditions and in particular on the access charges to be paid for cross-network communications. It is shown that, under a variety of alternative assumptions on the type of retail competition, higher access charges tend to lead to higher retail prices (the access charge has to be paid for each communication terminating on another network, and thus increases the marginal cost of communications perceived by operators and induces them to raise their final prices; even if the total access charges paid to each other by the operators perfectly balance *in fine*, they

still affect the operators' incentives to raise prices). Hence, unconstrained negotiations on (possibly reciprocal) access charges are likely to generate undesirable outcomes (the operators indeed have an incentive to agree on 'high enough' access mark-ups in order to induce themselves to set monopoly retail prices; moreover, a non-cooperative determination of access charges may well lead to even worse outcomes, i.e., to even higher access and retail prices, due to standard double marginalisation problems).

Standard regulatory interventions such as the efficient component pricing rule may moreover not be very effective or even desirable in such two-way bottleneck situations. The interpretation of ECPR, incidentally, is not obvious in such situations. In particular, the notion of 'bottleneck' varies according to whether an *ex post* or an *ex ante* perspective is adopted: once consumers have joined a network, both the originating and terminating ends are bottlenecks, whereas *ex ante*, the complementary segments are (possibly imperfectly) competitive. Indeed, a first inspection suggests that ECPR may soften price competition. (For example, if the operators first negotiate the access charges before deciding on their retail prices, a high access charge can serve under ECPR as a commitment to charge a high retail price. And if the operators set their access and retail prices simultaneously, then under ECPR a network cannot undercut and gain market share without providing its competitor with a windfall gain on access, which again tends to soften retail price competition.)

Second, it is often suggested that, during a transition period toward competition, characterised by the presence of a large, well-established dominant operator, entrants may be handicapped by the incumbent's reluctance to provide access to its network on a reciprocal basis and at a reasonable price (entry deterrence concern). This again is validated by a careful analysis of competition between, say, a well-established telecom operator, with full coverage, and another operator, which initially has no coverage and must thus incur an investment cost related to the coverage it is planning to offer. First, if the entrant's coverage is small, the incumbent network has an incentive to refuse interconnection (or to delay indefinitely the conclusion of an agreement), since in the absence of interconnection it can corner the market at a profitable price. Second, if interconnection is mandated but each operator is left free to set its access charge, the incumbent has an incentive to set its own access charge at a prohibitive

level, as this constitutes a standard ‘raising rival’s cost’ strategy. This suggests that imposing reciprocity in the setting of access charges may be a particularly good idea when entry or coverage is at stake, that is, when the industry is still in an immature phase of network competition. However, reciprocity may not suffice to prevent entry deterrence. If for example the incumbent can charge different prices for calls depending on whether they terminate on its network or its competitor’s network, it can still effectively block entry even under a reciprocity rule by insisting on a high access charge, and by charging a high retail price for off-net communications. It thus discourages its own subscribers from calling the other network (a strategy the entrant cannot mimic if its coverage is small), and avoids paying high access charges to its rival. (Even if termination-based discrimination is banned, a reciprocity rule for the access charge may still not suffice to generate the best outcome, as the entrant has an incentive to underinvest in coverage (or other dimensions of quality) in order to soften competition: by leaving a large captive market to the incumbent, it indeed induces this operator to maintain high prices.)

### **C. Universal service obligations**

The universal service obligation (USO) occupies a prominent place in the policy debate regarding all major network industries in the EU. First, the report specifies the scope of its study and emphasises that it focuses on the specific notion of ‘universal service’; it is not meant to provide a systematic discussion of the much more general concept of ‘public service’ which has been traditionally a major building block of public policy in many European countries. Though related, public service and universal service are by no means synonyms. The two concepts have different contents, reflect different attitudes of the policy makers, and lead to very different policy implications.

The recognition of the need to correct or supplement the market mechanism for reasons of efficiency or equity has led, in many countries, to the emergence of the notion of ‘public service’, according to which some activities should be directly entrusted to public authorities (or at the very least be subject to a tight regulation). This tradition is particularly well-established in countries like France and Belgium, but many other European countries also have a long tradition of public service.

The scope of this notion of public service is rather large and, depending on the specific country, it may include a more or less significant number of goods, services and activities. In most instances, it includes the provision by public authorities (or administrations) of ‘essential’ public goods like national defence, police protection, justice, etc., but also sectors like education, health (care and insurance), social insurance, etc.

In many European countries, the notion of public service has traditionally been extended to include public intervention in some of the major network industries (such as electricity, telecommunications, postal services, railway transportation, etc.). As for education and healthcare, the rationale for such policies lies essentially in the specific characteristics of the underlying products and services. The goods produced by network industries involve various externalities (including problems of environmental protection). They often constitute essential inputs which are of crucial importance both for national security and for the growth perspectives of the economy. They often involve investment decisions that call for long-run (and even inter-generational) trade-offs which may not be accurately reflected in available markets. Last, but not least, they have strong redistributive implications which lead to the concern that market provision may not be sufficient to guarantee a fair and non-discriminatory access to these services for all income groups and in all locations of the country.

This philosophy of public service may be hard to reconcile with the liberalisation of network industries. To the extent that a public service requires the direct provision of the good by the public sector, its very nature is challenged by the introduction of competition. However, this does not necessarily mean that all the features of public service have to be given up if an industry is liberalised. Liberalisation can be accompanied by regulatory measures which aim at preserving some of the essential features of public service. The universal service obligation provides one example of such a regulatory policy.

The objective of Part two of the report is to clarify the (rather elusive) concept of universal service obligation and to see how liberalisation may lead to an evolution of the concept. Universal service obligations (USOs) are traditionally viewed as the obligation for an operator to offer specified or all services at a ‘good quality’ level, at ‘affordable rates’, and to ‘all’ users. The quota-

tion marks in the previous sentence refer to the fact that items enclosed by them are not well-defined. As a matter of fact, the USO concept is often made more precise by additional requirements such as uniform pricing across consumers or geographical areas. The drawback of such refinements of the concept is that they side-step the goals and focus on the instruments.

There are really two issues concerning USOs. First, what is their rationale? Second, how should they be implemented?

Concerning the *raison d'être* of USOs, the report makes a distinction between rationales that are relevant and those which seem minor, if not flawed. The report, as it should, takes a normative perspective on the matter. But it should not be forgotten that USOs are often the outcome of a political process and therefore are not necessarily motivated by social welfare considerations. On the consumer side, some pressure groups have been quite successful in obtaining subsidies that are hardly visible, and are strong proponents of the current system as they are concerned that these subsidies might be removed, were they to become more transparent. On the producer side, incumbents may use the existence of USOs as an excuse to stall liberalisation, and entrants may lobby in favour of strict restrictions on the incumbent in order to weaken its competitive position.

In the policy debate, USOs are often motivated by the existence of network externalities, on the basis that enabling more consumers to use a service benefits other consumers of that service in the presence of network externalities. The report argues that, as it stands, this argument has limited relevance. On the one hand, a number of USOs in network industries are not specifically designed to boost network externalities. On the other hand, operators may benefit if network externalities are exploited, to the extent that they are able to capture some of the increase in consumer surplus, as has been illustrated by a number of unregulated industries (software, VCRs, and so forth). Similarly, a 'public good' approach to USOs seems to require further argumentation.

The report therefore argues that the currently well-grounded rationale for USOs is the redistribution concern. This rationale can be further divided into the desire to help low-income groups and that of fostering the development of poor or underpopulated areas. But, although such redistributions are desirable, they do not

by themselves vindicate USOs. Redistribution to poor people a priori should be based on income redistribution and as much as possible avoid distorting relative prices. That is, it is often more efficient to provide the poor with income rather than to encourage over-consumption of specific goods and services. But, even taking for granted that income and wealth taxation are imperfect, so that it may be optimal to use taxes and subsidies on consumption goods in order to achieve redistribution, it must still be examined whether redistribution should take the form of subsidies on telecommunications or water or of subsidies on education, child care, healthcare or food. (Of course, such subsidies need not be inconsistent. In any case, one must pay attention to the consumption of poor people relative to the rest of the population in order to know whether the inefficiency attached to the subsidy indeed has a redistributive benefit counterpart). The same issues arise with regional redistribution. One must assess whether income transfers would not dominate redistribution through prices and whether the latter cannot have perverse effects (as may be the case for some transportation projects that speed up a region's decline). In short, the report argues that redistribution through prices has good foundations but that one should not take such manipulations of prices for granted, and thus that one should conduct thorough comparisons of alternative redistributive policies.

The report next turns to the (related) issue of implementation. For a given redistributive effort, one would want to choose the redistributive instruments that distort the economy the least. The report discusses a methodology for comparing dead-weight losses for alternative policies.

The report then focuses on the design of USOs in a liberalised environment. The first possibility is to impose both the USOs and their costs on a single operator. This option creates cream-skimming and entry by less efficient competitors. To provide the operator subject to the USO with a guaranteed source of financing to cover the cost of the USO, one can allocate it a monopoly ('reserved') segment. This policy however unduly restricts the 'tax base' from which the USO deficit must be covered.

The second option is to again impose the USO on an exogenously determined operator, but to ask all operators to contribute through 'taxes' on the 'industry' services. Taxes can be levied on retail services or on the



USO operator's access charges (other revenues like those stemming from the auctioning of licences can also be used to cover the USO deficit). Taxes on retail services and on access charges are quite similar, except in the presence of bypass opportunities or of variable proportions technologies, for which a tax on access distorts the rivals' choice of inputs. It should be emphasised that optimal taxes are not purely cost-based and must reflect the intensity of competition and the elasticities of demand in the various segments. (See the discussion on access pricing in Chapter IV above.) The report also discusses the properties of 'menus' such as those defined by 'pay-or-play'-type taxes.

Last, one may refrain from selecting a priori an operator for the provision of USO services. One can for instance organise an auction in which firms announce the compensation they demand to fulfil a prespecified obligation in a given area and the lowest bidder is selected. This method has attractive features, but also raises the standard issues of market power in the auction, and especially of the incentive for the operators to properly invest in infrastructure if their assets are later auctioned off. An alternative method consists in setting up a system of taxes and subsidies, but not specifying who is entitled to receive the subsidies. In such a system, all operators can serve target consumers or areas and be subsidised for the corresponding services.

## **D. Policy implications for European network industries**

We briefly review below the main policy implications of the above economic analysis for the future of European network industries. Part three of the report discusses those issues in further detail.

### **I. Telecommunications**

The main policy issue in the telecommunications industry lies in the determination of the modes of access and the access prices. Purely cost-based access faces a serious measurement problem. Imperfect cost-based prices may raise substantial difficulties, cream-skimming or price squeezes. This suggests that alternative, more demand-based access prices, for instance in the form of a global price cap, may be worth exploring. The main objectives should be to avoid inefficient duplication, to promote efficient entry and to allow for the recovery of sunk investments: in this respect, too high lease or access prices lead to the undesirable duplication of seg-

ments of the network and to inefficient bypass, and can also deter efficient entry, whereas too low lease or access prices generate inefficient entry and also create recovery problems.

The cost of USOs is also a much debated issue, which may be studied using engineering models of networks. Such models are however still in their infancy and have recently been (May 1997) deemed by the FCC to be too imprecise to be used directly in the determination of the access charges. Regarding its financing, it seems desirable to use taxes with the largest possible base rather than access charges. Some form of competition, through auctions for franchises subject to quality control, may also be explored.

### **II. Postal services**

This sector has some specificities, declining demand and high labour intensity, which imply that any policy resulting in a substantial loss of market share for the historic operator may not be accepted on political grounds. Hence one of the major issues relates to the nature, degree and speed of liberalisation.

Distribution is the segment of the industry which has most of the characteristics of a natural monopoly. Since, as shown by the Swedish and Finnish experiences, the historic operator will probably retain a dominant position in the market at least within a foreseeable future, the question of providing downstream access to its network may arise.

The last key issue concerns the design and the financing of the USO: currently, USO is financed through cross-subsidies and the grant of a monopoly protection for some products, which artificially reduce the 'tax base'; alternatives, such as access surcharges or the constitution of a USO fund might be explored.

### **III. Electricity**

One of the main issues is to ensure a competitive behaviour in generation. In particular, the English pool system has been criticised for leading to excessively high prices: with a few generators, the potential for price manipulations seems high; on the other hand, excessive competition may make it difficult for generators to recover their fixed costs.

A second issue concerns the ownership and the pricing of the usage of the high-voltage grid. The determination

of adequate nodal pricing and the coordination of investments in generation and in transportation are the main issues there.

On the distribution side, the main issue is to design efficient access charges so that large customers can directly deal with generators without impairing economic efficiency. Demand management, that is the coordination of suppliers' investments in extra capacities and of customers' investments in substitute sources of energy, has also to be taken into account.

Lastly, USOs do not appear to pose a problem (since they are in general restricted to household users, who are not allowed to choose their supplier).

#### IV. Natural gas

One of the main issues in the natural gas industry concerns the vertical structure of the industry. Should the bottleneck owner face line-of-business restrictions, that is, should there be a divestiture of the incumbent's supply or transportation units, and possibly of seasonal storage facilities? The merit of such a proposition would be to establish a level-playing field among firms if none of them is integrated forward into transportation, while its drawback would be to possibly generate a lack of coordination between supply and transportation investments.

In the short term, the main objective is to select the lowest cost suppliers. Hence, as for electricity, prices should differ across nodes so as to reflect transportation constraints and to provide customers and suppliers with the right signals. In the case of a vertically-integrated transportation system, efficient access prices must be designed so as to encourage efficient entry while inducing the vertically integrated operator to invest and maintain the network properly and not to try to deny rivals' access to the bottleneck. In the long term, the main issue is the coordination of investments between suppliers, pipeline operators and large customers. A potentially interesting incentive mechanism for the transportation owner would consist in the design of a price cap on a basket of nodal prices.

#### V. Water

The regulation of local networks first requires the choice of its operator, possibly through auctions, and then poses the classic trade-off between the need to give this operator adequate incentives for efficiency (which

would favour a price cap approach) and the desire to extract the monopoly rents (which would tend to favour a rate-of-return or cost-plus approach). The fact that the network infrastructures are underground however creates additional difficulties in the evaluation of their quality, in particular when they are old. Moreover, if the regulator chooses to use concession contracts, one must assess *ex ante* the extent of the investments needed to replace the used elements of the network and to maintain them, and the monitoring of these necessary projects is also a difficult task.

The global management of the resource raises a substantial problem due to pollution. This is complicated by the fact that what is observed is not the level of pollution of individual agents but only that of a group of agents. Moreover, as water runs along rivers that often cross several regions, the issue of the appropriate geographic region under jurisdiction and the coordination among the regulatory authorities of the various regions are important questions.

#### VI. Urban transport

There is increasing support for private participation in providing urban transit services. However, choosing the adequate (de-)regulation is still debated. Like all others, this industry does not seem to constitute a contestable market, and it is thus not possible to rely on potential competition. Similarly, a system of urban transit organised around the competition of several firms is hardly conceivable, due in particular to the need of coordinating different types of vehicles and services in the use of the same set of transport nodes, and to the fact that some services are more profitable than others. Competitive tendering seems to be an efficient way to allocate the rights of access.

Another set of issues relate to the need to improve inter-modal passenger transport and the environmental impact and the congestion costs.

#### VII. Air transport

The main issue in the air transport industry concerns the allocation of airport slots. The liberalisation of the industry in Europe can only bear fruit if entry on the main routes is not impeded by the difficulty in obtaining landing slots. The use of auction-type mechanisms should be explored, along with consistent regulation of the usage rights of slots.

Another important issue is the need for an adequate competition policy. The US experience shows that competitive pressure in itself is not sufficient to prevent airlines from exploiting their monopoly positions. For example, the prices of flights from USA hub airports dominated by one airline are higher than the prices of similar flights from airports where several lines compete. Among others, frequent flyer programmes, code sharing, and the management of computer reservation systems should be examined as potential barriers to entry and/or tools for collusion.

Competition may not be workable for small routes, where increasing returns to scale are important. It may then be desirable to subsidise the service of those routes, using a compensatory fund as allowed by Council Regulation (EEC) No 2408/92.

### **VIII. Railway transport**

The main issue in the railway transport industry is how to introduce a degree of competition in the sector in order to improve the efficiency of the companies and adapt supply to demand, not only in quantity but also in quality. In particular, which authority should be in charge of the regulation, and how are the access prices to the network and the allocation of the slots on the different portions of the network to be determined?

Currently, access pricing (when available) is entirely cost-based, which will be the source of systematic disputes between operators if competition becomes effective.

Another set of issues relate to the interoperability of European domestic rail networks: although this may primarily seem a technical problem, the lack of interoperability can also be used strategically to deter entry in each other's markets.

Regarding freight transport services, it may be desirable to dedicate parts of networks to freight services only, in order to promote 'trans-European rail freeways'. This raises again the issue of interoperability, but also the question of intermodality. Solving the latter issue requires some technological solution, such as the use of containers that could be indifferently transported by ships, trucks and trains, but those 'trans-European rail freeways' would also benefit from the emergence of 'one-stop shops' which would organise the services on the whole corridor.

A last issue relates to the transition period, where existing, overstaffed national operators may not be competitive (which may also create political obstacles to the liberalisation). An interesting approach has been adopted in Germany for DBAG: on the one hand, a new structure was created and the current employees of the existing railways companies could choose between being directly employed by DBAG under private sector labour conditions or joining this new structure with their civil servant status; on the other hand, DBAG can use the services of these employees under the conditions of private sector, the federal government subsidising the difference. With this solution, DBAG does not support non competitive cost conditions while its former employees preserve their social advantages.

# PART ONE

## Deregulation of network industries and the role of the public authorities

### I. Network industries and network goods

#### Introduction

Are network industries extraordinary economic entities, characterised by features so specific that they escape traditional analysis? Or are they just the meeting place of complex but classical economic problems? The answer to this question is not purely academic. There is a strong policy concern in the idea that a network industry is analytically inextricable because it cannot be dismantled in any way without losing its intrinsic qualities. The observation that compatibility and interconnection produce substantial benefits to the consumer might suggest a rationale for monopolisation, vertical integration and regulation. In practice however, various segments can be served by non-integrated firms and exposed to some form of competition. For instance, in the water, gas or electricity industries, there is no evidence that efficiency is impaired when upstream stages of the production process (collection, extraction, generation, etc.) and the downstream stages (distribution, supply) are fulfilled by different players.

Modern economic analysis is now well equipped for the analysis of network industries. New tools have recently been developed that shed light on alternative policies regarding investment, production, pricing and regulation of these industries.

To get some insight into the economic problems raised by large network industries, we begin by giving definitions and illustrations of the concept of a network (Section 1). Then, we exhibit purely technological costs and benefits of network activities in Section 2. We focus on gains from horizontal and vertical integration which are essential in network activities even if they are not specific to these activities. The third section is devoted to a discussion of the economic advantages and drawbacks specifically due to a network organisation: density economies on the one hand and club and congestion externalities on the other hand. In Section 4 we give some intuitions on how networks can tolerate vari-

able doses of static and dynamic competition without losing their efficiency. Finally in Section 5 we present a typology of network industries based on the former characteristics and helpful for policy-oriented decisions.

#### 1. The concept of network

##### 1.1. Definitions

There is a strong heterogeneity in the population of the so-called network industries and the frontier with other types of industries is somewhat fuzzy. That electricity, telecommunications or postal services must be studied as networks is not questionable. In contrast, food retail, newspapers or insurance services, although they possess some similar features (e.g. multi-point distribution), are not often called networks.

In its simplest economic definition, a network is a set of points (or nodes) and interconnecting lines (or edges) organised with the object of transmitting flows of energy (electricity, heat), information (sound, data, pictures) or material (water, freight, passengers, etc.). Each point can be an originating node from which the flow is emitted, a terminating node, that is a node receiving the flow, or a node that plays an intermediary role of transmission, storage, amplification, coordination, dispatching and so on. Some networks are one-way, like gas, cable TV and water delivery while others are two-way, such as passengers transportation or telephone.

The essence of a network is that (almost) every pair of nodes can be linked by more than one line. Consequently, the path between an initial node and a terminal node is generally not unique, allowing important organisational advantages, but also some costly drawbacks when the flows cannot be perfectly controlled. This characteristic of multiple potential links explains why many economists enlarge the notion of network to informal organisations, such as the set of the users of VHS video-recorders or the users of PC-type



computers. Their common feature is that flows of products or services can be easily transmitted between any two points in the network, with or without the help of a material infrastructure.

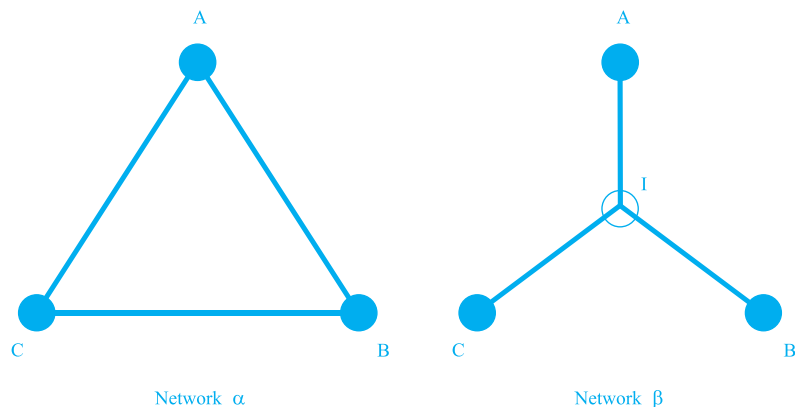
Given the flow transfers to be performed, the best network is the one that minimises total costs. Consequently, the design of a network results from a

trade-off between building costs and operating costs. When building costs are very high as compared with operating costs, the best network contains few lines, which can require the creation of interconnection nodes. On the contrary, when operating costs are dominant, the best solution is a dense network with most final nodes directly connected to each other, without any intermediary (see Box 1).

**Box 1: Dense versus minimal network**

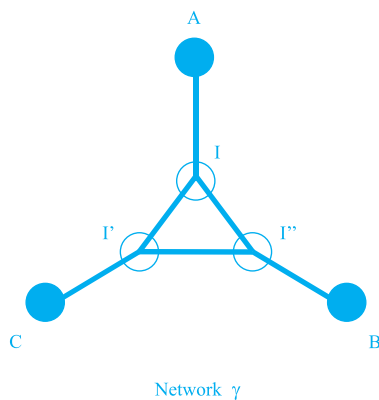
Suppose that three agents A, B and C are to be connected within a network. The first solution is a ‘dense network’ like network  $\alpha$  that connects directly any couple of the three agents. It is clearly very costly in terms of infrastructure investment. On the contrary, network  $\beta$

is a ‘minimal network’. In this second type of connection, one needs shorter lines but it is necessary to create an intermediary node I, which any flow will have to cross, incurring operating costs for screening and dispatching.



Most networks have an intermediary configuration like network  $\gamma$ , reflecting the trade-off between operating costs and building costs. Of course, a final node can be

used as an intermediary node for storing, translating or rerouting the flow.



## 1.2. Some illustrations

At this very simple level of definition, we can give examples of network configurations and begin to present some of the economic and political problems they raise. For air transport of passengers, the essential determinants in the configuration of commercial networks are the intensity of the administrative regulation and the development of computerisation in airport management, together with the economic cost of buying (or renting) and operating planes. Since aerial routes are free and since the companies of air transport can lease almost anything (airport slots, aeroplanes, crew, reservation systems, etc.), the design of their commercial network is actually very sensitive to the public authorisations or obligations to operate airlines. While most domestic airline networks have long been close to the model of configuration  $\alpha$  in Box 1 under regulated access, in countries where competition is at work, airline companies have adopted a 'hub-and-spoke' system, that is a minimal configuration  $\beta$ . In most cases, the intermediary point chosen (the hub) was an airport that already existed: for instance point I was located in A which resulted in the disappearance of the direct line between B and C (see also Box 3). Nevertheless, the new design of the networks has resulted in costly arrangements of airports and of company organisations to create hubs, which show that the suppression of lines entails high connection costs.

An illustration of a dense network  $\alpha$  is given by the citizen-band addicts who can call each other within a given area using free airwaves. Clearly, the building cost of lines is zero and they do not need any intermediary node for dispatching their calls. Each member can send a message to all others and can receive their messages, provided they are simultaneously present on the same wavelength. If two members want to communicate privately, they have to agree on unusual particular timetables and wavelengths, or to adopt a cipher.

In railway transportation, both building and operating costs are very high. Competition by aeroplanes, cars and trucks has resulted in a loss of density of the rail networks as compared with the first half of the century. The construction of new railnets is highly expensive because of the price of land and the requirements of fixed equipment. This and the maintenance costs explain configurations with a minimised length of lines. But the shorter the total length of lines, the longer the average duration of travel, which increases the operating cost. As the designers of the network have to take into account the total cost, the final layout is always a compromise. As an illustra-

tion, network  $\gamma$  in Box 1 is not very different from the map of the high speed train planned for Texas in 1992, with A = Dallas, B = Houston and C = San Antonio.

The identification of the final points of a network (and consequently its size) is exogenous or endogenous depending on the nature of the flow transported. When the product is not storable, as is the case for electricity, the network necessarily connects the production nodes to the final consumption nodes. On the contrary, for storable goods like letters, the network can be arbitrarily shortened upstream as well as downstream, each user sending or receiving mail at home, or at collective boxes, or at the central post-office. The same is true for water, but because of its permanent use and of public health requirements, the best solution is a direct connection of the drinkable water pipes to the taps of the final user and of the final user's draining system to the sanitation network, except if a cheaper local alternative is possible (e.g. sinking a well). Similarly for gas, each user has the choice between the connection to the distribution network and repeated purchases of gas cylinders. In areas with a very low density of population, the connection costs to a distribution network are too high and the use of cylinders will be preferred.

Also, the design and the operation of a network strongly depend on the homogeneity of the product. A kilowatt-hour of electricity, a cubic metre of drinkable water or a cubic metre of domestic gas are completely standardised products. When they are injected at one point of the network, their destination does not matter since each unit is a perfect substitute for any other unit, which highly simplifies the dispatching. Clearly, this cannot be true for personal letters or telephone service, or for passenger transportation. In the latter cases, each unit injected into the network is identified from the very beginning and no substitution is allowed. Sorting, bulking and unbulking are essential and costly stages in the operation of the network. The identification of units circulating through the network is also essential for measuring and billing individual consumption. For electricity, gas or water this can only be done at the very final level of consumption while telephone meters can be grouped at a central level since calls are totally identified by the numbers of the agents on both sides of the line.

The path followed by a unit of product in the network is more or less under the control of the operators. Actually, in almost all networks an analogue of the so-called Kirchhoff laws is at work, that is flows have a

natural tendency to take the path of least resistance. When the flow is circulating slowly enough (e.g. water or gas) its path can be approximately controlled by switching on or off the commutation nodes. But in the telecommunications and electricity industries, the high speed of flows limits the permanent control of the paths. In network  $\alpha$  of Box 1 for instance, if agent A wants to send energy to agent B at a time when C is connected, a part of the total injection will necessarily pass in transit through C. This transit can result in inconvenience and even in damages for agent C.

Most networks can be interpreted as a set of interconnected sub-networks. For instance, network  $\gamma$  can be seen as the four-point network {C, I', I'', B} connected to the two-point network {A, I}. Historically, most national or international networks have been developed by progressive interconnection of small local organisations. Interconnections raise problems of compatibility in the characteristics of the flowing product and in the operating procedures. Interconnecting nodes and lines are critical components of the entire architecture. The possibility to disconnect very rapidly some subsets from the global infrastructure is of great importance for any network. Actually, such a possibility is essential for the integrity and survival of some activities, for example in the distribution of drinkable water to prevent contamination and in the electricity sector to prevent total collapse, like the one that resulted in France on 19 December, 1978 from an unusually high demand.

### **1.3. Operators and users**

A wide variety of agents are concerned by networks: designers, builders, owners, users, customers, operators, regulators. Several questions result from this variety and from the nature of network activities.

First, as network infrastructure is generally installed for a very long time, the private or public nature of designers and builders is not neutral in so far as the decision horizon is shorter for private institutions. Short-sighted decisions will result in under-investment both quantitatively and qualitatively. The same kind of problem appears for the use of the geographical space. Absent any private incentives to take account of the environmental damages, one can think that the infrastructure installed will not be exactly the same when initiated by a public or by a private agent <sup>(1)</sup>. This is one of the rea-

sons why the utilisation of public property is to be tightly monitored. For instance, the public authority can keep the property rights on the network industry and exclusively concede its operation to a private firm after a selection process, like in the water activity or for urban transportation. The operator's behaviour is restricted by quality or environmental qualifications as well as by universal service obligations. The main drawback of the concession system is that since the government keeps the control of access to the industry, when the operation requires sunk investments the firm will not invest at the optimal level fearing a potential expropriation. When investment can be recovered, one can use alternative ways to regulate private operators. For instance in telecommunications, the operator is the owner of the infrastructure and, as a residual claimant, his investment can be supposed to be more efficient <sup>(2)</sup>.

The infrastructure is used by two types of agents: service providers and service customers. In simple activities, like the first telephone networks, only final users are present. They are directly in contact with the infrastructure operator who interconnects them on a demand basis and charges them correspondingly. When the complexity of the operation increases, new agents can be necessary to satisfy and to stimulate demand. The supply of enhanced services (inquiries and interconnections in telecommunications, reservation and catering in passenger transportation, etc) can be performed by the infrastructure operator or by new distinct agents or by both. If separate suppliers of services are present, their activity is strongly dependent on the infrastructure operator. Either a provider of final services or not, the operator is very powerful since he controls the access to final users. To prevent risks of abuse, explicit rules and tariffs of access should be established on an efficiency basis.

The ownership of the infrastructure is a sensitive problem in any network activity since the infrastructure clearly appears to be a 'public good', that is a good that can be used simultaneously and/or successively by several agents (and often by a large number of agents) without any loss in the quality of use. Additionally, the use of the network by one agent can create indirect beneficial or detrimental effects on the other agents. The control of and payment for these externalities need the presence of a 'super-agent' able to rationalise the behaviour of small independent

<sup>(1)</sup> It does not mean that public firms are necessarily environmentally friendly. Some are big polluters.

<sup>(2)</sup> Unless a tight price regulation is expected, so that the private operator anticipates an expropriation of profits through very low prices. Then, although the owner, he is a residual claimant on nothing, which drives his investment incentives to zero.

users. This agent can be a private, a public or a mixed entity but it needs some collective concern to handle the maintenance and development of the infrastructure.

## 2. Costs and benefits from integration

Because of their complexity, one could think that networks cannot be analysed with classical economic methods. This sometimes alleged inability would be the mere reflection of the intrinsic integrity of networks, with the political conclusion that large integrated firms are the only suitable solution for an efficient organisation of network activities. Actually, as we show in this section, modern economic analysis has numerous high performance tools in hand and keeps on increasing its toolbox. Some features of networks are already well-known as they are shared with many other economic activities, for instance economies of scale or economies of scope. Others are more recent fields of knowledge and deserve more research, such as the gains from rerouting or the economies of clubs. Thanks to these instruments, when the need for coordination of activities is so intense that it entails some degree of integration, economic analysis is able to find it out without dogmatism. Moreover technical integration and economic integration are not synonymous and economic integration is not necessarily the best way to reach efficient coordination.

### 2.1. Economies of scale and economies of scope

In any economic activity, economies of scale become apparent in the decrease of the average cost of production when output is expanded. Economies of scale result from large fixed costs and/or weakly increasing variable costs. In all the networks where a physical and/or logical infrastructure <sup>(1)</sup> is essential, strong economies of scale are at work since once the equipment is installed, any additional unit of output lowers the fixed cost per unit of product. This unit-cost decrease is effective up to the technical limit of the equipment, but some gains from 'squeeze' are sometimes still available for a given saturated equipment as we will see later (Section 3.1 below).

Now, at the moment the infrastructure is designed and installed, one can rely on the so-called 'surface/volume effect' to lower the unit cost when output is an increasing function of the plant volume while total cost increases with the envelope of this volume (see Box 2). This effect leads to the installation of very large equip-

ment to obtain very small average costs. It is at work in the design of equipment for storage and transportation of oil, gas, freight and parcel post. The surface/volume effect is also essential in the evolution of passengers' transportation by plane (wide-bodied jets), bus or train. For train, the size is obviously limited by the size of bridges and tunnels as well as by the width of railways. Also note that when the size of the equipment is increased, the operating costs increase less than proportionally. For example, in the transport of freight, the crew of planes or boats is almost independent of the size of the planes or boats, as well as it takes one person to drive a truck, whatever the truck.

The variable-cost component of economies of scale is less obvious in network activities. First because the definition of the activity can be imprecise due to vertical integration (see Section 2.2 below). For instance, generation costs are part of the total cost in an integrated power system while manufacturing costs are obviously not included in the cost of a freight transportation network. For this reason, the successive stages of the activity should be disconnected, at least from the accounting point of view, before any attempt to quantify the economies of scale. A second reason is that in most networks, output has an essential geographical characteristic. For instance in freight transportation, output is measured in tonnes-kilometres. Consequently, output can increase because of an increase in the quantity of commodity transported, because of an increase in the delivery area or because of both, with obvious divergent effects on operating costs. With a given equipment, operation costs are more or less proportional to the quantity produced but increasing more than proportionally with the distance, except in telecommunications.

The more standardised the product, the stronger the economies of scale. But for heterogeneous products, economies of scope can still be at work. In networks, when the infrastructure of storage and transport is not totally dedicated to one specific activity, there still remain possibilities of gains from diversification of activities. The reason is that the infrastructure has a nature of public good which means that less than the stand-alone cost is incurred when one adds a new activity to the other activities.

Definitely, a pipeline cannot transport water if it is currently used to transport gas, or an oil tanker cannot be occasionally devoted to the transport of food products. But on the other hand, in a warehouse one can store a

<sup>(1)</sup> By logical infrastructure, we mean the whole operating system of the network.

wide variety of commodities and a distribution network can deliver very heterogeneous parcels, including passengers, with minor additional costs. For instance, some planes have two lives: during the day they transport passengers and at night, seats are removed and they are used as postal planes.

Telecommunications are the industry where recent changes in technologies have dramatically focused on the gains from economies of scope. Thanks to the use of optical fibres and of digital signals, it has now become indifferent to transport data, sounds or pictures.

Consequently, a telephone company is a good candidate for database transfers or a cable TV network can enter the telephone market at a reasonable cost <sup>(1)</sup>. Moreover, when wires are to be used, the control or ownership of a part of the public property is essential. This explains that some potential challengers in telecommunications (broadly defined so as to include transfers of data), such as power, gas, water and rail utilities are companies with access to the public property.

<sup>(1)</sup> The cable operator will have to upgrade its one-way network to make it a two-way network

### **Box 2: The gains from large scale operation**

Let  $TC(q) = FC + VC(q)$  denote the total cost of production  $TC$  with a fixed component  $FC$  and a component  $VC$  increasing with the level of output  $q$ . There exist economies of scale when the ratio  $TC/q$  is a decreasing function of  $q$ . The higher the output, the lower its total unit cost. In most networks, the costs of infrastructure are very high so that the unit fixed cost  $FC/q$  is an essential part of the total average cost and economies of scale are important on this side of the saturation threshold of the equipment. For  $VC/q$  to decrease, the variable cost has to increase less than proportionally to output, which can be explained by standardisation and learning.

The surface/volume effect is a strong determinant of the large scale of many industrial, storage or transportation equipment. In many industries, building and maintenance costs are directly dependent on the surface of the physical envelope of the equipment (walls and roof of warehouses, tubular steel of pipelines, hull of tankers, etc.) while the output can be measured by the volume of the equipment (commodities stored, gas injected, oil transported, etc.). As surface increases with the square of the equipment scale while volume increases with its cube, the ratio surface:volume, that is the average cost, is a decreasing function of the scale. This effect leads to the building of very large equipment, limited by the resistance of materials and by environmental considerations.

Stochastic gains result from the decreasing probability that independent events occur simultaneously when the number of events increase. Then the reliability of a production or a transmission node increases with the number of substitute pieces of equipment installed while the maintenance cost (spare parts, repairers) does not increase in proportion. In telecommunications, for a given quality of service (a given probability that a call can be transmitted), the ratio of the entrant traffic to the number of network gears is increasing (Erlang law).

Economies of scope denote the ability of firms to produce several goods at a total cost smaller than the sum of the costs of these activities isolated one from the other. If we consider the production of quantities  $q_a$  of good  $a$  and  $q_b$  of good  $b$ , one simple definition of scope economies is  $C(q_a, q_b) < C(q_a, 0) + C(0, q_b)$  where the costs on the right-hand side of the inequality represent the stand-alone costs. These economies result from the ability to use some pieces of equipment and a part of the workforce for producing goods or services that are not too different.

An extreme case of scope economies is represented by inputs that possess the nature of 'public goods', i.e. that are not destroyed by use, like informational inputs. For instance, the data base made of the clients of a bank can be used without any additional cost to propose insurance services. Then economies of scope are a strong incentive for the diversification of activities.

Vertical integration refers to the performance within a single firm of a number of successive distinct operations needed to produce a given commodity. From a public point of view, vertical integration can be justified in so far as it allows to decrease costs for producing the final product. This decrease can proceed from technical complementarities, for instance through a better coordination between the successive stages of a production process. Also, the social gains can result from the removal of private profit margins obtained by intermediary firms with some market power. But upstream vertical integration is also a device for increasing the control a firm already has on its suppliers, clients and competitors. By controlling the delivery of certain strategic inputs, a firm can win a dominant position on its final market. In network activities, the most obvious example is the control of the access to transport infrastructure by a firm that is also a user of this infrastructure.



## 2.2. Vertical integration

Any production process requires a sequence of elementary operations that can be performed either within the same firm or by independent firms. A vertically-integrated firm controls all the successive operations until the final consumer. For instance in the electricity industry, vertically-integrated firms providing generation, transport and distribution are common. The drawback of this arrangement is the lack of competitive incentive at the intermediary stages of the production process. The upstream units are too far from the end of the process to be influenced by market reactions and they are certain that their output, even if its quality is low, will be accepted by (inside) users. For this reason their level of effort will probably be far from efficient <sup>(1)</sup>. But vertical integration has strong advantages in terms of access control and in terms of coordination.

Rapidity and reliability are necessary qualities in the operation of networks. When an activity is not vertically integrated, each part of the production process is under the control of separate entities with potentially divergent objectives <sup>(2)</sup>. This can result in a very low performance as compared with an operation under the control of only one supervisor. Suppose a passenger takes a plane to go from A to C and there is a stop at an intermediary point B. When segments AB and BC are operated by distinct carriers, it is less likely that arrival at and departure from B will take place at the same terminal, or that the timetables will be compatible, or that luggage will not need a new registration, etc. In all the transport activities, vertical integration can facilitate the resolution of problems at the nodes, particularly the transshipment of cargo in multimodal transport. In electricity, the need for coordination is obvious as the flow is not storable and moves at the speed of light, but technical coordination can be performed by an independent firm.

The social drawback of vertical integration in network activities is the abuse of power in the control of access to infrastructure. In any economic activity, the control of an essential input is a major cause for upstream integration, and conversely the owner of an essential input has a strong incentive to integrate the downstream activity. Thanks to this control, a firm becomes responsible for the quality of the input in question and for the

regularity of deliveries. But it can also take unfair advantage on its use which is bad from a collective point of view.

In network activities, it appears that the essential input is the transmission infrastructure. For example in electricity, it includes the transport infrastructure for high-voltage transmission and the distribution lines for the electricity under low-voltage. As the good or service has to be delivered to every final user and as it would be too costly to install more than one delivery system, the one who controls the transmission system clearly controls the access of any upstream provider to the final market. And if the infrastructure owner is also a final-service provider, a fair competition is problematic <sup>(3)</sup>. For this reason, vertical integration is not necessarily the best economic answer to the need for technical vertical coordination. For example, in most countries, airline companies are not the owners of airport installations. On the other hand, some have got the control of reservation networks which also are essential inputs for this activity. In the British experiment for rail traffic, train operators will be separated from the rail and station operators.

For competitors, the alternative is to bypass the infrastructure, either by duplicating a new one or by developing a new technology. Setting-up a new infrastructure is bad on pure technological grounds because economies of scale from each network will not be totally exploited. This occurred at the beginning of the service-network era with the duplication of gas first, then electricity networks in most countries. It still exists in North America for some railroads and in some USA towns for cable TV. In telecommunications, wireless technologies allow new entries without having recourse, at least partially, to the installed infrastructure. TV transmission by satellite is a complete bypass of cable operators.

## 3. Density, externalities and interconnection

In this section, we present economic features that are more specific to network activities and that are to be taken into account in the appraisal of the performance

<sup>(1)</sup> To foster efficiency, one can try to mimic market incentives using complex reward schemes.

<sup>(2)</sup> Objectives are also divergent within an integrated organisation, but it is the very rationale of the organisation to limit the degree of divergence.

<sup>(3)</sup> In the US gas industry, before the 1992 Order, entrants used to complain against the transportation firms for interrupting their service when the operators needed the pipeline capacity for their own firm sales' service.

of operators. First we introduce the costs and gains from 'compression' that are present in all the network activities even if their importance is not the same everywhere. Then, we show how the positive and negative externalities created by the use of the network can change the behaviour of the agents and result in a wide variety of alternative equilibrium configurations.

### **3.1. The economies of density**

Since one of the objectives of network industries is to interconnect isolated agents, one can think that the more numerous the members of the network, the more efficient the industry because the infrastructure cost is divided between a larger number of members. Nevertheless, networks are strongly dependent on geographical characteristics and the number of agents can be meaningless without any reference to the area affected. In the same way, since networks are used for transferring flows, one could say that their performance is increasing with the global quantity of flow processed. Actually, a given flow passes in transit through lines at the same time as other flows or in a specified order, crosses other lines that transport other flows, is stored, transformed, re-routed at certain nodes where other flows are processed, etc. Consequently, a measure of performance is highly dependent on the physical and organisational characteristics of each network, more precisely on its ability to transfer several flows simultaneously or sequentially with a high quality of service. So, it appears that density (the density of connected agents as well as the density of goods or services processed) is a feature indispensable to estimate the performance of any network.

Consider first the geographical density of agents to be interrelated within a network. Clearly the cost for connecting a given number of people is increasing with any measure of the distance between them. Pipes, wires, ways or fibres are expensive to produce and to install. Moreover, operating costs are increasing with distance because of the equipment and the quantity of energy needed for transport or emission. If agents are scattered too much, one single interconnected network can be a very poor arrangement. It would be less expensive to organise small local, mostly non-connected networks.

In the case of water for instance, most distribution networks are organised under pure geographical considerations using gravitational forces. The essential constraints are the natural availability of water at each point

(waterway, ground water, etc.) and the slope between the taking nodes and the supply nodes. Historically, urbanisation has increased the need for networks of drinkable water and sanitation, as Roman remains show. The present outcome is a large number of small networks without any permanent interconnection within a super-network, except when there is a structural shortage in a region.

Consider now the case of electricity. It can be produced with very different technologies, from small to vast, so that power plants can be installed almost anywhere. Physically, one can imagine electricity without any network, each user generating just enough to satisfy his own needs. However, networks have been installed from the very beginning of the electricity era. Large-scale economies are obviously part of the explanation but the main reason was the need for light in high-density urban zones. For interconnections of local networks, the main drawback is that power losses are increasing with distance and are proportional to the square of the current. But the use of alternating current and high-voltage transmission allows to limit these losses because alternating current may be easily converted to higher or lower voltages by means of transformers and because the greater the voltage on the line, the less the current. Additionally, on a given line the flow of power can go in any direction. Consequently, networking appears as a low-cost solution for providing a non-storable good to people with non-simultaneous consumption except for remote and/or non permanent locations. The additional cost of large networks is that they need an organisation able to coordinate numerous decentralised decision nodes. In the electricity industry, dispatchers have been playing a central role from the very beginning.

High density of users does not entail only advantages. Crowding can be the source of serious drawbacks mainly in transport activities. When people want to trade more material commodities or decide to travel more, the higher the density and the more likely their lines cross each other. Costly works are necessary to prevent congestion of lines and of storing nodes as well as collisions at crossing points. For instance to regulate the traffic at cross-roads, some cheap 'software' solution can improve the performance (traffic lights, no U-turn or no left turn signals). But if traffic is very dense, 'hardware' changes will be necessary, for instance road broadening or subway digging. An intermediary option lies in arranging the cross-roads with a roundabout.

To increase the flow of commodities, passengers, information or power through an installed network at low cost is the dream of all operators. In some activities like gas or water this can result from higher pressure, at the cost of more fuel and sometimes new pipes. In electricity, better performance is obtained thanks to the use of very high-voltage transportation, more efficient transformers and improved insulation of lines. During the last two decades, outstanding progress in the transport of goods has been made due to the standardisation of containers which allow to shorten the delay for transshipment. Also, the development of several types of coding (bar code, postal code) has greatly facilitated the regulation of flows and inventories of goods.

For the next decade, one can expect that the industry most affected by these 'squeeze gains' will be telecommunications. As far back as 1875, Edison's quadruplex boosted the development of telegraph because it allowed the simultaneous use of the same line by two operators on one end of a line and two operators on the other end. So Edison succeeded in multiplying by four the capacity of each telegraphic line. Similarly, the use of digital signals instead of electromagnetic signals allows a fast growth in all the forms of communication. Digital data can be generated directly in a binary code by a computer or can be produced from a voice or visual signal by an encoding process. These coded data can then be easily compressed before they are transmitted, on the condition that the receiver possesses the device to decode them. The result is, for example, that on a given TV channel, one can transport eight times more digital programmes than analogical programmes. Due to this increased density in the transport of signals, there is an enlarged variety of choice for any individual user (helped by computers) and/or a cheap possibility of

transforming any one-way line into a two-way line (like telephone). Then, all economic activities that need interactivity between agents could be dramatically changed. Many fields will be and already are concerned by digitalisation: telephone and TV but also many forms of entertainment and education as well as many service sectors (banking) and distance shopping.

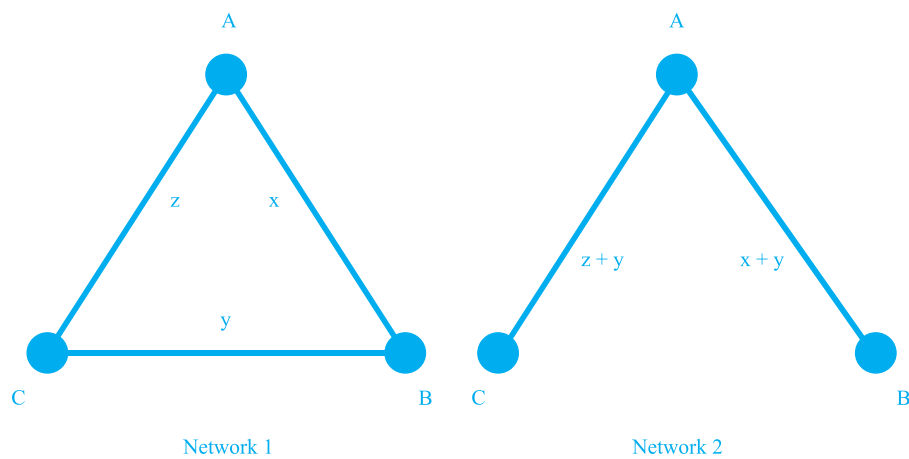
Of course squeezing is not completely safe. First because a network with a high degree of compression is likely to have a low quality of service so that clients can be discouraged from using it. But one can think that the operators, or at least the regulator of the sector, will limit the level of compression. Second because compression is obtained at the cost of a low redundancy which means less security. Take the example of a firm operating a water network. When a new group of customers appears, the operator has several possibilities to supply them. First a direct independent connection to a take node, which is not the best solution in a high-density population area. Nevertheless this solution is very secure since the risks of cuts are divided between two independent infrastructures. Second a simple connection to the existing network, which is very cheap but can provoke important drops in pressure resulting in a high dissatisfaction of old and new customers. Third, a connection to the existing network joined with an increase in the injection pressure. Clearly now the risk is that old pipes cannot support this new constraint and that leakage increases, lowering the overall performance. A fourth alternative is to let the users install and operate their own equipment for example to pump up water and store it in tanks. Its main advantage is to fit the individual needs. Also, pressure is not increased in all the pipes. But it has the social drawbacks of being more costly than a centralised 'booster' and of decreasing pressure in other parts of the network.



### Box 3: The gains from rerouting

In a network, the optimal way from one point to another is not necessarily the shortest path. When designing the infrastructure as well as when operating it, one has to take account of the traffic on all the neighbouring edges

and through all the neighbouring nodes. As several possible paths between two points exist, one portion of or the whole traffic can be diverted to minimise total cost.



Consider for instance the traffic forecasts  $x$ ,  $y$  and  $z$  between nodes A and B, B and C, C and A respectively. If  $D_{ac}(z + y) + D_{ab}(x + y) < D_{ab}(x) + D_{bc}(y) + D_{ac}(z)$  where  $D_{ij}(\cdot)$  stands for the total cost function for traffic between  $i$  and  $j$ , it is worth constraining all the flows between B and C to transit through node A. The solution

will be minimal like Network 2 instead of the dense Network 1. The drastic solution is to never build line CB.

Actually the solution can be intermediary, that is line CB is installed but a portion of traffic between C and B is transferred through point A. Consider the following illustration:

Line	AB	BC	CA	Total
<b>Traffic forecast</b>	$x$	$y$	$z$	
night	1	3	1	5
day	2	1	2	5
<b>Capacity</b>				
independent solutions	2	3	2	7
optimised solution	2	2	2	6

On lines AB and AC, traffic is lower at night than during the day and conversely on line BC. In a sub-optimal solution, one installs on each line a capacity equal to the maximal flow to transfer, which means a total capacity of 7 for the whole infrastructure. A better solution consists

in installing less capacity on BC and at night to transfer one unit through the lines AB and AC. The resulting network is dense during the day and almost dense during the night.

### 3.2. Clubs and networks

In a network, the interdependency of agents generates strong externalities, that is effects of the decisions by each member on the level of satisfaction of other members. These externalities can be positive, which means that an action by a member increases the utility of the others: we will refer to this type as 'club externalities'. On the contrary when they affect negatively the utility of other agents we will speak of 'congestion externalities'. In many cases, the size and composition of a network are the result of a trade-off between club benefits and congestion costs.

Positive externalities enhance the utility from being a member of a network and the larger the network, the stronger the club effect. In networks where final users are the main stimulating persons of the service, like telephone or electronic mail, club effects are direct, which means that the gain from being a member is directly increased by the entry of new persons. There exist indirect club effects when one benefits from the large number of members only through some equipment that possesses the nature to be a public good. For instance, the holders of a specific credit card are better off when they become more numerous because more automatic teller-machines are installed and more shops accept this mean of payment. The same kind of externality exists in informal networks like the users of computers and video recorders: the more numerous they are using one specific electronic device, the richer the variety of programs or films the producers will make available <sup>(1)</sup>.

At a given access price, when the composition of the network changes qualitatively or quantitatively, some members can be induced to leave while some non-members would like to be registered. This means that the number and/or identity of those who wish to be in depend on the number and/or identity of those who are in. An equilibrium is reached when demand is compatible with itself, that is when the number of actual members (and possibly their identity) is exactly the same as the number (and identity) of the candidates (see Box 4). But because of the club effect, at a given price there generally exist several equilibria.

Now, all the potential equilibria are not identical. Some are stable because any slight exogenous shock like accidental entry or exit will be counterbalanced by a spontaneous exit or entry letting the final configuration unchanged. Others are quite unstable: a small initial shock in one direction or the other develops into a 'snowball effect'. For instance, as some members leave, the club loses its attractiveness and others will follow out. If this second wave of resignation is larger than the first one, it is the beginning of a positive feedback which can end only with the complete vanishing of the network. The smallest unstable equilibrium is the 'critical size' <sup>(2)</sup>. It is essential to identify it before trying to launch a new network. If the operator is not able to reach this minimal equilibrium in one jump, he will never succeed. Many examples of this failure can be found in the telecommunications industry. In France, a service of data transfer by radiowave named Mobipac will be interrupted at the end of 1996 only three years after its beginning. In the same way, Bi-Bop, a mobile telephone service will soon disappear because of the disaffection of users. In England the 'Telepoint' mobile phone started in 1989 and disappeared in 1992.

Of course, the equilibrium size of a network is also highly dependent on the state of the infrastructure at each moment. If too many people want to use a telephone network at the same time, they can provoke a complete breakdown. Why do networks usually break down? If we except a simultaneous shock on several essential parts (flood, storm, etc), the main reason is that a line is broken and the other lines cannot stand the (automatic or semi-controlled) rearrangement of flows. For instance, thermal constraints become suddenly too high in an electric grid, such as happened in California at the beginning of August 1996. Alternatively, an isolated problem in one site can spread through the network and can result in a failure of the whole system if no safeguard device has been installed. In computer networks for instance, if disconnection systems are not installed, a small initial incident on one computer can provoke the chain interruption of all the programs, computer after computer. For example, on 7 August 1996 America On Line, a gateway provider to Internet, suffered a complete collapse during the whole day.

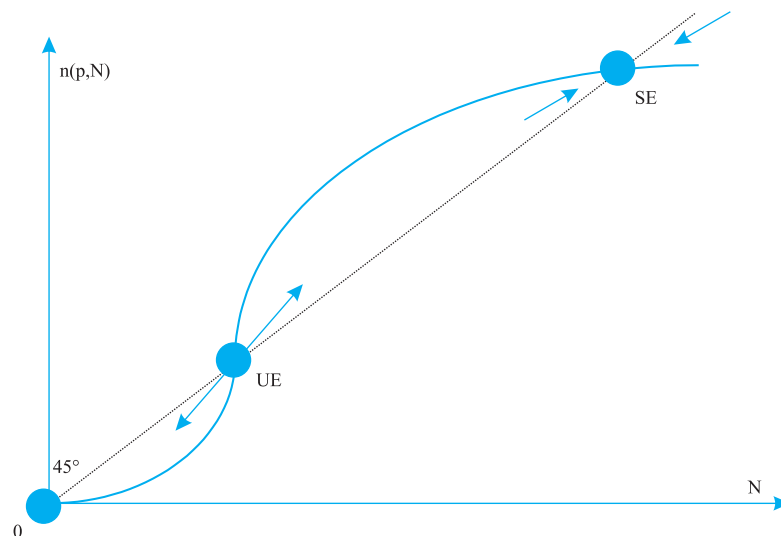
<sup>(1)</sup> Actually, these 'indirect club-externalities' result from economies of scale and/or of scope: when the number of customers increases the network operator can decrease its unit cost and/or propose to sell a wider variety of services. This effect is at work in any industry.

<sup>(2)</sup> It should be stressed at this stage that the equilibria depend on prices, the access price as well as the price of use. So actually there is not one unique 'critical size' but one for each level of the overall price. The one we refer to here is the critical size corresponding to the lowest price the operator can charge without losing money.

#### Box 4: Network externalities and multiple equilibria

In order to illustrate positive externalities that appear in networks, denote  $n$  the number of people who would like to be members of a club and  $N$  the number of actual members. If  $p$  is the entry fee,  $n(p, N)$  is decreasing in  $p$ . But it is likely to be increasing in  $N$  because a large club means a high quality of service and many opportunities to meet people and to establish relations. For a given  $p$ , if  $n(p, N) > N$  like in the central part of the figure below, the actual network contains fewer members than the number of those who would like to be members. A new entrant transforms the size of the network from  $N$  to  $N + 1$ ,

creating an additional externality. If this externality is very strong, it creates a 'snowball effect' that creates a large network. When this positive feedback is exhausted, one reaches an equilibrium — that is a size  $N$  — such that  $n(p, N) = N$ . For such a value of  $N$ , no outside agent wants to enter and no inside agent wants to exit. In the figure, equilibrium is reached when the curve  $n(p, N)$  crosses the 45° line. Conversely, when  $n(p, N) < N$  like on the left and right parts of the figure, a process of downsizing is at work.



Depending on the shape of function  $n(\dots)$  and on the value of  $p$ , the 'membership curve' can cross the 45° line only for  $N = 0$  or many times, which means that there can exist a large number of equilibria. Some like SE are stable equilibria: if a member leaves for any exogenous reason, he is replaced by a new one. Others like UE are unstable because if one member leaves, several others follow him and if a new member subscribes, he incites many others to do the same. For a given price, the highest stable equilibrium is the saturation size of the network. The smallest unstable equilibrium is the critical size. If the operator is not able to reach this threshold, the network cannot exist at this price level. Put in different

words, for this price the equilibrium size will be the smallest stable one, that is  $N = 0$ .

One can find a simple example of repeated network externalities in public or private lotteries. The number of participants  $n$  is a decreasing function of the ticket price  $p$  and an increasing function of the jackpot  $J$ . Assuming that all the money collected is included in the jackpot, there is an equilibrium when  $p \times n(p, J) = J$ . Clearly, several equilibria are possible. A small  $J$  will attract few people which means little money to be included in the prize. On the contrary, a large jackpot should be attractive enough to collect the money necessary to pay it.

## 4. Networks and competition

When we speak of competition in network activities, one must specify whether competition takes place *ex ante* or *ex post*. The reason is that even in the activities where economies of technical integration are so strong that a monopoly is the most-efficient social solution (e.g. very-high-voltage-electricity transportation), *ex ante* competition can still be used as a regulation device <sup>(1)</sup>. For instance, a natural-monopoly position can be allocated as a temporary franchise through an auction process, like the sealed-bid auctioning of the water concession in Buenos Aires. If the process is opened to all specialised firms and designed to be collusion-proof, it will result in the allocation of the franchise to the most efficient candidate. And if the incumbent is threatened to lose its position at the end of the concession period when a new bidding procedure is opened, it will be stimulated by strong incentives to manage efficiently the industry it has in charge.

In this section, we focus on *ex post* actual competition, that is competition between two or more firms already active in a market. We first show how network specificities modify the model of competition used in traditional industrial organisation (Section 4.1). Then we examine the problem of dynamic competition between network firms (Section 4.2).

### 4.1. Competition under network externalities

Probably the most specific characteristic of competition between networks operators is the very large number of potential equilibrium configurations for a given set of competing firms. This multiplicity results from the club effect described formerly in Box 4. Because of the club effect, the attractiveness of a given network is closely dependent on what each potential customer thinks about the total number (or the identity) of customers in each competing network. The point is that each firm offers a service the quality of which is endogenously given by the participation of clients. And it is the expected quality that matters more than the true quality, if any. This expectation effect is a decisive determinant in the dynamics of competition between network operators as we will see in the next section. In any case, competitors cannot survive if they do not reach the critical size cor-

responding to the price they have chosen (see Box 4). This 'market threshold' is an additional limitation to the number of active operators in a given network industry, the first limitation being the usual technological threshold: if firms are too numerous they cannot pay back their operating costs. For all observers of the English mobile-phone experience (Telepoint, 1989–92), the failure resulted from the inability of the four franchisees to set up a base large enough for their network to be attractive to new customers. On the contrary, at the beginning of the French Minitel experimentation, France Telecom had chosen to distribute the terminals for free despite its monopolist position. Later, after the critical threshold was cleared, France Telecom organised the sale of the terminals in its local agencies.

A second key feature of network competition is that it can be transformed into some form of cooperation by decisions of compatibility, sometimes unilateral decisions. When the utility of goods depends on the number of their users, a decision of compatibility between several products or services dramatically changes the size of individual customer-networks. For instance, when the two French credit-card networks decided to make their cards compatible in 1984, the utility of being a user was instantaneously increased. Compatibility can be achieved by standardisation. The most striking example is the electricity industry where there is a long tradition of cooperation at the national and international levels, with frequent power exchanges through interconnections. The reason why producers may prefer not to standardise is that standardisation makes products more homogenous. Consequently it can enhance competition and reduce the profit margin of sellers. For each producer, the trade-off is between a small weakly-competitive market (few clients multiplied by a high mark-up) and a large highly-competitive market (many clients multiplied by a low mark-up). Standardisation can alternatively be achieved through the use of gateways. In many cases, it is a cheaper solution and in some cases, it can be decided unilaterally. For example in the activities of freight or passengers transport, an operator can prefer to use buses or trucks on given segments of his own network or his competitors' networks rather than to install railways or to operate airlines. Multimodal platforms that many regions are installing to improve the interconnection between their transport networks are gateways that allow to take advantage of complementarities of heterogeneous systems. But they also stimulate competition on some parts of the networks.

<sup>(1)</sup> The 'natural regulation' of industries with natural-monopoly characteristics through potential competition is presented in Chapter II: Government intervention in network industries.

From the social point of view, standards and gateways are not equivalent because of their effect on the product or service variety. While achieving compatibility through gateways increase the possibility of choice of the customers, on the contrary standardisation reduces the variety offered on the market.

Pricing is an essential piece of the competition among firms in any market. This in particular holds when these firms are network operators. When they propose very similar services, price competition is fierce because price is the only element that differentiates sellers in the opinion of buyers. The best known example is the competition between airline companies on a given destination (even with railway companies when the distance is short enough). To avoid this competition, operators try to make their services heterogeneous, by improving the provided quality or by specialising on particular market segments. But obtaining these segments can be a source of competition by itself. For example, all airline companies would like to get the most profitable take-off and landing slots, early in the morning and late in the afternoon. The allocation of these slots often remains traditional: the oldest firms are privileged by 'grandfathering' practices. But in the USA, some airport authorities have begun to allocate them by means of auctions, and in some countries there exists a grey market where slots are reallocated.

In networks where a physical or an electronic connection is needed before the service can be consumed, pricing competition is more complicated because the service providers can combine access fee and service fee in such a way that it is very difficult for a given user to know where the cheapest offer is. For instance, for French mobile telephones, three operators are currently proposing free or expensive connection (to get the telephone set), monthly or quarterly subscriptions (to get the right to use the telephone set) and linear or multilinear prices for the utilisation. The use of complex multi-part tariffs has several advantages for the sellers. First, as we have just mentioned, it makes the tariff grid obscure for the users so that it relaxes competition. Second, it is a way to win the loyalty of customers. When customers have to pay a two-part tariff with a low variable part, the cost of switching to a new seller before the expiration of the contract is represented by the high fixed-part. This is an essential feature of the competition between the satellite and cable operators for TV programming as well as between telephone operators. Third, it is a legal method to discriminate

between customers. By proposing to all users several combinations of fixed and variable prices, the operators lead each customer to select an average price different from the price paid by others. This discrimination is authorised despite the usual uniform-price obligations, because the same grid is proposed to everybody and discrimination results from a process of self selection.

As we have seen formerly, some pieces of network infrastructures are not totally devoted to a specific activity. When an infrastructure admits alternative uses, the operator can change very quickly its activity or he can diversify, which means to be present simultaneously on several markets. In the case of quick moves from one market to another, competition is more intense since the number of sellers can change very rapidly. For example, an airline company can easily modify the allocation of its planes and crews if it appears that some lines are more profitable than others or if the access to some protected lines is now authorised. In freight transport, if trucks are not dedicated to a given product, the producers can adapt their behaviour to sudden variations in demand. Therefore, alternative but not simultaneous use increases competition. On the contrary, when a given piece of equipment can be simultaneously used for several activities, the operator can run it as a mechanism of cross-subsidisation. Then there is a possibility of unfair competition on some markets. For instance, with the same telecommunication infrastructure it is now possible to send sound, pictures or data. Consequently, cable TV operators who possess a strong position for delivering entertainment or sport programs can collect resources from their TV activity and compete fiercely in the provision of telephone services.

Finally, we can evoke the access problem that will be developed further in this report. How to organise competition around an essential infrastructure efficiently? In the activities where a given equipment is necessary to deliver a good or service to final customers, if it is too costly to duplicate this equipment the firm that controls it has the ability to discriminate between the service providers. The classic example is the transport and distribution infrastructure for electricity generators. With this power in hands, the infrastructure manager can spoil the competition between generators by mimicking the behaviour of a vertically integrated firm. The problem can be worse if the firm that operates the infrastructure is also allowed to sell the final service like in the British telephone sector. The operator is then simultaneously a provider and a competitor for the final service

producers. In any case, when there exists a natural monopoly position somewhere in the successive stages of an industry, it is difficult to imagine true competition in other stages of this industry without an economic regulation of the monopolised segment.

#### 4.2. Investment and dynamic competition

The necessity to reach a critical size to survive and the endogenous quality of networks modify the usual dynamics of competition between firms. First, because the initial decisions can be critical for the success or the failure of a given service provider. When the decision to subscribe to a provider is costly, customers want to be sure of the quality of each provider before they take their decision, because their expenses are unrecoverable. Then each competitor tries to persuade the potential users that his network will be the best one. If enough customers can be persuaded that a given network will be the best one, it will be so because a large number decide to subscribe and they enhance the quality of the network. Then, in these activities, competition is fierce at the earlier stage. Firms have to spend large amounts of money in advertising and in investment to gain the credibility of clients, and to obtain their subscription, they have to propose low introductory prices. These three initial weapons (large investment, costly advertising and low initial prices) were used by JVC (Matsushita) to win the domestic-video battle against Sony and Philips in the early 1980s. On the contrary, in England, none of the first franchised operators of Telepoint (Callpoint, Zonephone and Phonepoint) followed this strategy and they all missed the set-up of a mobile telephone network in London. Note that in some cases, the identity of users is more important than their number. Consequently, the network operators have an incentive to quickly attract some particular pioneers, for instance large firms for a data-exchange network.

Because of switching costs, users are locked-in and competition between the operators that survive to the set-up phase is softer. Then profits are high for survivors, and these expected high profits are a stimulus for competing at the initial stage. When connection is not needed, the sellers can still impair the mobility of customers through some premium like the frequent-flyer system of the airline carriers.

For the aforementioned reasons, network firms have a strong incentive to invest in infrastructure at the very beginning of their activity, trying to create a momentum in their favour. When the industry becomes mature and

each surviving firm has constructed a large installed base, as competition gets weaker the incentives to invest and to maintain the quality of service decrease. Then technological competition is not very intensive in mature network industries, unless an innovation creates some dramatic change like wireless telephone or data compression did. Without a drastic technological improvement, the entry of a challenger into a mature network industry is almost impossible. It would need a strong price-cut to compensate the club advantage of the incumbents.

The dynamic regulation of the industry is an essential determinant of the investment policy of competitors. For example, in sectors such as mobile telephone where access is restricted, the licensees adapt their investment decisions both to the present and to the expected intensity of regulation. They can have an incentive to over-invest if they expect an increase in the number of licenses for the use of the same technology. The aim of the large initial investment is to create a large installed base, hoping that the induced club externality will deter future entry. In the same way, the present regulation of prices restricts the future benefits or the future revenues from investment according to the type of regulation. Then, it directs the nature and the size of investments.

### 5. A typology of network industries

The wide variety of network industries suggests several alternative typologies depending on the objectives of the classification. For instance for technical or legal reasons, the analytic key can be the nature of the flow transmitted. Thus we could oppose energy networks (electricity, heat) to material networks (passengers, water, gas, postal services) and to information networks (telecom, TV, electronic money). Alternatively, for medical, legal or/and ethical reasons, one can make a distinction between the transport of persons (air, train, bus) on the one hand and all the other types of networks on the other hand.

From an economic policy point of view, the problem is to know whether competition can be used as a regulatory mechanism in some network activities or administrative regulation is necessary. A broader discussion of the justification for public intervention is presented in the next chapter. Here we only consider the efficiency argument. With this restricted approach, competition should be impeded only when it could limit the benefi-



cial effects from vertical and/or horizontal integration or sectorial agreements.

As compared with other industries, it results from our previous developments that the main distinctive features of a network are first the existence of strong externalities (positive club externalities and negative congestion externalities) and, second, their node-edge structure. The cost of installing and operating such an infrastructure as well as the loss of club benefits can be strong reasons to blockade the entry of any competitor.

In Table 15 we try to summarise the main characteristics of the industries analysed in the first part using the two keys mentioned above: the node-edge structure and

the existence of externalities. We have classified the industries in three groups. First we think that local networks deserve a special treatment because of the narrowness and the specificity of the markets concerned. In this first category, we find the water sector as well as the urban-public-transport activities (upt). Other industries not studied in the first part, for example waste collection or urban heating, belong to this category as well. The other industries, with a national or an international concern are classified according to their 'one-way' or 'two-way' nature. The motivation for this distinction is that the nodes, which are essential points of control in any network, are more complicated and harder to manage when they are transit points in two directions.

*Table 15*

**Main characteristics of network industries**

	One way electricity, gas	Trains	Two-way telecom, postal service, airlines	Local networks water, urban public transport (upt)	Notes
<b>Lines</b>					
Infrastructure	no duplication	no duplication	duplication possibilities competition (postal service, airlines) imperfect competition (telecom)	no duplication	duplication can be justified by: — yardstick competition — differentiation — dynamics — security — pollution and congestion in urban areas
Substitutes	partial substitution	strong substitution	low substitutability for telecom (but Internet ...) strong for postal services (fax)	no substitutes for water strong substitutes for upt	
<b>Nodes</b>					
	gas is storable at nodes electricity not storable	stations are monopolies	airports are local monopolies	depollution units have minimal Optimal Scale	terminal nodes and intermediary nodes can have a very different status with respect to competition
Upstream	competition in production for electricity gas storage can be competitive	multimodal competition in dispatching	multimodal competition	upt: multimodal competition	
Downstream	distribution monopolies	multimodal competition	telecommunications and postal services are switching from monopoly to multitechnological competition	water: competition in treatment natural monopoly for water distribution	
<b>Externalities</b>					
	Kirchhoff law	congestion	club externalities in telecom	congestion	— indirect club externalities everywhere
	need for dispatching	need for coordination	congestion	quality variations	— problem of compatibility

Electricity and gas use one-way networks where it appears that the critical element is the transport infrastructure. While the generation equipment can be duplicated on different scales and with various technologies, the duplication of electric lines would be too costly as compared with the gains from competition. The same is true in the gas industry with the storage installation on the one hand and the pipelines on the other hand. This is also true for water distribution but the problem is to be solved at the local level.

For railways, stations (nodes) and rails (edges) could hardly be duplicated at low cost. But the rolling equipment is likely to be owned and operated by distinct agents and there exists a strong intermodal competition. In air transport, airports play the same role as stations for trains: they are obvious natural monopolies. But the lines can be opened to a large number of competitors, provided their flight programs are coordinated by an independent agency. In all these transport activities, the other transport modes are both substitutes and complements. The interconnection between the different modes enhance the efficiency of the whole transportation network.

Telecommunications are the sector most affected by recent technological changes. While some decades ago, it was unthinkable to connect telephone users without a network of wires and manually-operated switchboards, which means a non-duplicable infrastructure, nowadays

the technical progress both in hardware and software allows users to chose their long distance operator. Within the next decade, they will probably be able to bypass the local loop. Therefore, the telecommunications industry is the one that can be widely opened to the regulation by market mechanisms.

Finally, from an economic policy point of view, it appears that the key problem is to know whether, in a given network, the nodes and the lines can or cannot be duplicated at a reasonable cost. And if duplication is too costly, can nodes and lines be bypassed at a reasonable cost by potential entrants? This approach allows to identify bottlenecks, if any, and to decide when an administrative regulation is to be installed. Table 16 presents seven network industries focusing on their ability to support competition, monopoly, or a mix of free market and planned organisation.

This type of classification is very sensitive to the definition of each network. For instance, an electric network is not usually restricted to the transportation and distribution of power. It is defined as including the generation nodes. On the contrary, a railway network is limited to the transportation between stations. The upstream grouping of passengers and freight as well as the downstream degrouping are realised by other modes of transportation. Also, in terms of possibility of duplication, it appears that the notion of node, without additional qualification, is not totally pertinent. Indeed, while entry

Table 16

### Bottlenecks, bypass and duplication possibilities

	Entry mode		Lines and nodes		Exit mode
<b>Electricity</b>	generation	transformation	transport (dispatching)	transformation	local distribution
<b>Gas</b>	generation	storage	pipeline (dispatching)	storage	local distribution
<b>Water</b>	pumping, treatment		distribution/collect (basin regulation)		treatment-dumping
<b>Postal service</b>	clearance	sorting	transport (international coordination)	sorting	local distribution
<b>Telecom</b>	local loop	switch	long distance (allocation of audio-frequencies + coordination)	switch	local loop
<b>Airlines</b>	intermodal	airport	transport (traffic control)	airport	intermodal
<b>Railways</b>	intermodal	station	transport (traffic control)	station	intermodal

Key: **Competitive activity**; *mixed activity*; monopolistic activity.



nodes can be easily identified in the power industry (generation stations), they are not in the postal service sector unless we accept that each household or firm can be considered as an entry node. For this reason, it is more convenient to discuss the possibility to introduce some form of competition into networks in terms of entry and exit modes rather than in terms of nodes. This is how we have labelled the first and the third columns in the preceding table.

With this presentation, it clearly appears that the entry mode is able to accept high degrees of competition in

most industries. On the contrary, the network activity per se is intrinsically a natural monopoly position in at least one of its essential parts: nodes for airlines, railways and postal services; lines for water and perhaps for gas, both for electricity. The telecommunications sector is quite atypical in this set. Finally, with the exception of transport (because of intermodality) and telecommunications (by anticipation to the bypass of the local loop), the exit mode is also a natural monopoly position, that is an activity where some central regulation is needed, particularly restrictions in terms of universal service obligations.

## II. Government intervention in network industries

### Introduction

The predominant view in policy and academic circles is that the age of large public or regulated monopolies has lived, and that, after a period of transition, we will witness the emergence of network industries where competition will reign and where there will be very minimal regulation. This view raises two sets of important questions. First, up to the 1970s, there was in the same circles a consensus that in those very industries technological conditions imposed the presence of only one firm, which had to be controlled by the government. This view was held by nearly all policy analysts, economists and politicians. Two hypotheses can explain this collective change of mind. Technological conditions have changed and may have lowered ‘transaction costs’, especially computing and communication costs, enough to make competition desirable where it was not. Alternatively, some argue that the progress of economics has shown that past opinions were wrong, and among other failures that they did not take into account seriously enough government failures. Policy analysis has been revised accordingly. In any case, understanding those changes has much more than academic interest: if we do not want to repeat past mistakes, we must understand why they were made.

The second important set of questions that the new consensus raises concerns the roles that governments will still have to play in the future of network industries. The transition to more competition has seen extremely detailed intervention by regulators. Is this a temporary phenomenon? Is there any reason to believe that in the long run, the role of government will be different in network industries than in other industries?

We will answer those questions by analysing the traditional arguments for government intervention in the economy, and asking ourselves to what degree they apply to network industries. It is obvious that there will be government intervention in these industries, as there

is in all of economic life. We will therefore focus more precisely on the following questions:

‘What are, if any, the specific aspects of network industry that, according to economic theory, warrant greater regulation than for other industries? Will the form of that involvement be different than in these other industries?’

The answer to the cost benefit-analysis implicit in this question — what are the comparative costs and benefits of government intervention — will depend on the method by which the government intervenes. We therefore begin by analysing the different methods that are available, stressing the newer market-based techniques such as auctions. We will then study in turn the two main arguments for government intervention, correcting market failures and redistributing income. The potential benefits of government intervention must be weighed against its costs, and we discuss them in Section 4.

### 1. Modes of government intervention

Before analysing the reasons why government should intervene, we review in this section the main techniques that it can use to promote its aims.

#### 1.1. Legislate and enforce the law

Governments intervene in the simplest possible form through legislation, that is mandating or forbidding some types of behaviour. Typically, legislation will also provide for penalties for parties who break it. Enforcement takes two forms:

- private parties can ask for redress if another party has caused them harm by breaking the law;
- governments put in place institutions whose role is to monitor compliance and to ask for penalties when it is lacking.

There are a few special features of network industries with respect to the law.

- Those industries are often very capital intensive, and require long run commitments from the parties to some of the contracts. For instance, the construction of a new airport requires very heavy complementary investments in means of access. The purchase of aeroplanes is planned many years ahead. The legal system must be able to handle efficiently the very complex resulting contracts, by giving the parties to these contracts easy access to legal redress if they have been wronged.
- Giving the importance of many network industries for national welfare, governments often give to an individual or an agency a special role in verifying that firms abide by the law. This role can be held by the same individual(s) who are also in charge of regulation.

## **1.2. Tax and subsidise**

Governments collect taxes, levied on income of individuals or firms, on transactions, on wealth, and subsidise some activities. As we will see in the discussion of income distribution below and in Part two, in which universal service is discussed, these taxes and subsidies are often implicit, when they order a firm to fund a decrease in the price of a product or service through an increase in the price of another one.

## **1.3. Sell, rent and buy**

### **1.3.1. The problem**

#### **1.3.1.1. Examples**

The right to use airways to transmit phone calls or messages belongs to the government, or, more precisely, to the national community. Airwaves are used by private parties, telephone companies and their clients. Because the spectrum is limited, it is not a feasible solution to allow free access to the airwaves, and some mechanism must be implemented to choose among potential users.

Runways in airports can only accommodate a certain number of planes every hour. In very busy airports, the demand for slots is larger than the supply, at least at peak hours. Some rules must be used in order to ration this scarce good. In the EU, these rules form a grey area, but recent discussions around the BA–AA (British Airways–American Airlines) alliance, or recent mergers

of airlines show that implicit or explicit rights to slots are extremely valuable.

In the recent reforms of the British electricity sector, the grid needs to obtain electricity from suppliers. How should this be done? The solution that has been chosen is a daily auction, but one could think of long-run or medium-run contracts, with an attached price schedule.

In the water industry, local networks are often, and more and more often, run by private firms. The relevant governments need to choose among the candidate firms, and to offer a contract that induce the winner both to provide water efficiently and to maintain the very expensive capital. In this case, the government buys the service ‘managing the network’.

It should be stressed that network industries and public services are not the only important examples of government selling and buying. There are the obvious cases of procurement, for military or civilian needs, but other cases create problems that are closer to those discussed in this documents. For instance, environmental policy is basically the allocation of rights to use a common resource, the capacity of nature to absorb pollution, and market-based instruments are sometimes appropriate (see, for instance, the trading of SO<sub>2</sub> rights in the US).

The aim of this section is to describe the techniques that can be used in these types of situations, and to evaluate their merits and failures.

#### **1.3.1.2. Problems in the contracts**

In order to discuss the issues, there is one useful straw man. Consider the case of the choice of a company to manage a water network. Under the, very unrealistic, conditions where:

- (i) the costs of the firms are known to the government;
- (ii) the demand of the consumers is known to all parties;
- (iii) firms do not collude;
- (iv) there is no difficulty in writing a contract that describes unambiguously the responsibilities of all parties;

the optimal contract is easy to find. The government chooses the lowest cost firm, pays it just enough to compensate it for its costs, and contracts for the services that maximise social welfare.

In reality, these conditions that we have assumed are not met. We will review first the difficulties choosing a firm, and then the difficulties in writing a contract with the chosen firm.

### *(1) Choosing a firm*

#### *(a) Unknown costs*

When the government buys a service, it knows the cost of the firms less well than the firms do. This creates two difficulties.

1. Efficiency requires that the lowest cost firm be chosen. Generally, auctions solve that problem well as the firm with the lowest cost proposes the lowest price.
2. Firms have an incentive to overstate their costs. In the case of auctions, they will do so by proposing bids that are substantially higher than their costs. The experience of the privatisation of the regional electricity companies (RECs) in the UK, shows that this can yield very large profits. There is no perfect answer to this problem, but it can be mitigated by careful design of the mechanism used for choosing the firm, and by increasing the competition at that stage.

#### *(b) Unknown demand*

More difficult issues can arise when firms have better information than the government on the demand for the services of the network. There are two different sets of problems.

1. Some firms might be better managers and therefore the demand that they will face will be higher than the demand faced by their competitors. This problem is similar to the problem of unknown costs, and can be solved, in part, by paying the winner in proportion to revenue or, if revenue is insufficient, in proportion to the use of the network.
2. Some firms might have better information about general characteristics of the demand. This does

not imply that they should be chosen, as they might have higher costs than the others, but one would still like to have access to this information. Careful design of the mechanism used to select the firm can help on this issue.

#### *(c) Unknown willingness to pay*

The two points mentioned above, unknown costs and unknown demand are relevant for companies that manage services. When the issue is the right to use public resources, such as an electricity distribution network, the same type of problems arise but the main uncertainty is about the value for the firm of using this resource. Similar analyses apply in this case.

#### *(d) Collusion*

Often, the firms that compete for one contract compete for many others. Collusion is not difficult, and can be very profitable. The firms can take turns proposing the low bid, which is higher than would have been under competition. Alternatively, after a winner has been chosen side payments can be arranged, sometimes by hiring the competitors as subcontractors. We know relatively little about the best ways to hinder collusion. Rules that allow or require the buyer to exclude low bidders certainly favour it, as they make it easier to bring in line reluctant participants. It should be however stressed that collusion is not limited to market base instruments.

### *(2) Contractual difficulties*

Not only must one choose the appropriate firm, but also the contract must be designed in order to promote efficient use of resources. It is not feasible to discuss in depth all the relevant features, but a short description of the main difficulties can give an idea of the wide range of difficulties that have to be faced, although, of course, not all of them are relevant for all contracts.

#### *(a) Measuring the performance of the firm*

When the government is trying to hire the manager of a network, it must carefully choose the instruments by which the performance of the firm is measured. It should be remembered that one would really want to reward the firm in proportion to its contribution to the increase in the utility of consumers. This is impossible, but all other measures should be evaluated by how close they come to this ideal. Two techniques are generally used.

1. Sometimes, technical criteria are imposed: trains must arrive on time, water quality must meet certain criteria, service must not be interrupted more than a certain number of minutes per year. This is the appropriate approach when a few criteria provide a good proxy for quality. Too many criteria yield difficult problems of aggregation, while it is often difficult to measure dimensions of service such as the pleasantness of the staff of the firm. It is known that too much emphasis will then be put on the measurable criteria.
2. One can sometimes find a proxy for the value to the consumers of the service or good provided by the network, for instance when one pays the firm according to the revenues generated. This provides powerful incentives to serve well the 'marginal' consumers, those whose consumption is sensitive to the quality of the service, but consumers who do not have a choice of suppliers and whose demand is fixed will be less well served.

All these measures of output are actually proxies for what one would really want to measure: the effort made by the firm. In the design of the incentive schemes, one must take into account the fact that these measures are actually the result both of this effort and of some variables that it does not control.

The same type of problems arise when selling the right to use a common resource. First, the resource must be well-defined. One can give an airline a landing slot, and notice afterwards that it is systematically late. Second, the use of the resource is usually subject to the provision of a service of sufficient quality. This can be ensured, as in the case of the buying of a management service, either by bounds on measures of quality or by revenue sharing.

#### (b) Allocation of risk

In network industries, contracts between firms and governments typically have very long-term horizons; sometimes, perpetual property rights are allocated. There can be very large risks, and one should be aware of the fact that firms will ask for a risk premium. It is sometimes possible for the government to limit these risks or to bear some of them, and in so doing to improve social welfare.

For instance, some Chilean economists, Eduardo Engel, Ronald Fischer, and Alexander Galetovic (1997) have suggested that highway franchises should have variable time horizons. The franchise would hold up to the time at which total (discounted) revenues reach a certain level. This limits the risk linked to a bad estimate of demand, and they estimate that the increase in social welfare linked to this reform would be in the order of one fourth of the total cost of construction.

#### (c) Investment and maintenance

When the government hires a firm to manage a network, a delicate problem of investment and maintenance of the capital stock occurs. The contract with the firm generally has a fixed horizon, at which point a new firm is chosen. Either the capital has stayed property of the government, or it is transferred at the end of the contract. In any case, the firm computes the benefits of the installation of new capital or of maintenance of the installed stock in a different time frame from the government, and it will have incentives to underinvest and to provide too little maintenance.

This bias cannot be entirely corrected. It is less damaging if the contract is longer, but this has other costs. Measures of the value of the capital stock at the end of the contract, and provisions that guarantee that the firm will be compensated for its expenses help, but this can never be done perfectly.

Another problem linked to investment will be the tendency for firms to choose technologies in which they have a comparative advantage, in order to make it more costly to switch to another supplier in the following contract. They will also underinvest in formal information systems which would make it easier for competitors to take over. Again, contractual terms can mitigate these effects, but only partially.

#### (d) Modifying the terms of the contract

Ideally, a contract would never be renegotiated. It would include a list of possible contingencies, and describe how the terms are to be modified when they occur. This is impossible, and circumstances sometimes arise where the original contract is obviously not applicable. In this case renegotiation occurs. This renegotiation generally takes place in circumstances that are less favourable to the government, because there is less competition. The contract should therefore be as fool-

proof as possible, and the circumstances under which it is to be renegotiated well specified.

(e) Renegotiation and the 'hold up' problem

Because the firm's bargaining position is often stronger once the contract has been signed and it is difficult to call on a competitor, the firm will often try to force renegotiation, sometimes by not meeting all its obligations. Economists call this the 'hold up' problem. The contract should mitigate the risks of this occurring by specifying penalties that can be levied without too great a cost in case the contract is not fulfilled. Generally, it should make sure that any renegotiation takes place with an appropriate balance of power. This is very important, because the expectation that renegotiation can occur changes substantially the incentives of firms. For instance, a firm which has a comparative advantage in renegotiations will bid more aggressively; exactly the wrong effect from the view point of the government. Or investments will be postponed so that the firm has less at stake in the renegotiation process.

(f) The incentives of the government

Finally, it should not be forgotten that the contract also affects the incentives of the government. For instance, if a firm and the government share toll revenue for a highway, the government will have less incentives to fund a competing railroad than if the firm received all revenues. The firms will of course realise this and their bids and behaviour will be affected. In the interest of the government itself, the terms of the contract should therefore make sure that it has no incentive *ex post* to take actions that adversely affect the profitability of the firm.

### 1.3.2. Techniques

#### 1.3.2.1. What type of contracts?

The preceding section has discussed the main difficulties of writing a contract, and has provided hints of the solutions that can be used. Here, we would like to touch on one last point: how explicit should the contract be? Publicly-owned utilities have traditionally been managed within a very informal arrangement. The relevant member of government, or the relevant bureaucrats, would once in a while give instructions on priorities. Very often, the utility was in charge both of proposing public policy and implementing it. It was expected that because the managers have little financial incentives in

the profits of the firm, and were judged according to their contribution to social welfare, they would serve as well as they could the interests of the country as a whole. This pattern has been modified somewhat by the introduction of contracts between public utilities and governments, but these contracts are used more as declarations of intention than as binding documents. In particular, governments do not feel bound by their terms, especially when they have been signed by preceding governments run by another party. (The same problems arise with contracts with private firms, but the contracts are legally more difficult to ignore.)

As governments felt that public utilities were not providing the benefits that they were expecting, they have either tried to convert them independent publicly-owned agencies or turned to private firms. These require much more explicit contracts, in order to guarantee to shareholders profits that make it worthwhile to accept the agreement. Several techniques are available.

#### (1) Introducing competition

In some industries, the best type of control on the activities of the firm is the introduction of competition. Then, the basic contract between the firm and the government is general competition law.

#### (2) Contracting for specific services

Even if competition is generally feasible, there may exist some services that the government would like to be provided but which are not privately profitable, the provision of universal service is one leading example. In this case, a contract for the provision of this service can be written, and, depending on the industry, will be subject to the difficulties discussed above.

#### (3) Regulation

Regulation typically leaves much leeway to the regulator, and is open ended. In this case, the contract is not well specified. This has both benefits, one does not commit oneself to actions that would *ex post* prove to be non-optimal, and costs, the incentives of the regulated firm are not very powerful, because it knows that *ex post* the regulator can 'expropriate' at least some of the benefits of cost reduction.



#### *(4) Franchising*

In some industries, governments give the right to a firm to manage a sector for a limited, but sometimes long, time. A franchise can be national, for instance television licences, or local, for instance roads, water or railroad services. The cost of franchises is that the contract needs to be very explicit, and that it commits the government. Mistakes are very hard to correct, and franchise contracts require great technical expertise. The firms that specialise in industries where franchising is prevalent have strong incentives to learn how to use the rules to their best advantage, and to lobby for rules that favour them. It is imperative for governments to develop independent expertise.

##### *1.3.2.2. Choosing the firm(s)*

Once the type of contract is chosen, the government must choose the firm with which it will be signed.

#### *(1) Negotiations*

One strategy is to choose one or several firms with which to negotiate. If there are several firms, they will typically be asked to make proposals that outline their plans, and the government will choose among those plans, insisting on modifications. When there is only one firm, a similar process is used, but the starting point of the negotiation is more likely to be a proposal by the government. An advantage of this method is that it enables great flexibility to take into account information that is difficult to codify, such as effects on industrial or regional policy.

The same process can be used when the government is selling the right to use a resource. In this case again the negotiation will take into account other dimensions than price and how well pre-defined criteria are met. This is for instance the technique used by the French government to allocate the right to develop new mobile phone networks.

The flexibility of this technique is not only its main strength, but also its main weakness. Indeed, it gives opportunities for politicians and civil servants to pursue private aims.

#### *(2) Auctions*

The other available strategy is to organise an auction. In this case, the rules are clearly set in advance, and the terms of the contract are defined, except for some para-

meters that are left blank. In the simplest case, the only parameter left blank is the price, and the contract is given to the firm that offers the best price. In more complicated cases, several parameters are left blank and a rule is defined by which the winning firm is chosen. For instance, in an auction to run a regional railroad one could give weight to low fares, to high fees to the government, and to frequency of trains (other criteria, such as investment and maintenance being held fixed for all the bidders).

There are several important benefits to the use of auctions.

1. The most visible is that they can raise substantial revenues for the Treasury, and, contrary to taxation, do so without introducing too much distortion in the economy. This should be contrasted to the use of negotiations, which will induce substantial, but socially wasteful, expenditures in lobbying, public relations and other 'influence activities', what economists call 'rent seeking'.
2. More important, perhaps, from the viewpoint of social efficiency is the fact that auctions allocate the good to the bidder with the lowest cost or the highest 'willingness to pay', which will generally be the one who has found the most efficient use for the resources. The result of negotiations will depend on the benefits perceived by government negotiators of different uses for the resources, and because potential users present distorted evaluations of these benefits, these perceptions will often be inaccurate.
3. Auctions are generally perceived as being fair. Because the winner of the auction has competed with other potential firms, there is generally no windfall profit. Furthermore, because the criteria must be explicitly stated before the beginning of the auction, it leaves less room for favours to political friends or to members of an old school network.
4. Auctions seem to be able to function well in environments where markets are not efficient. Indeed, when coordination problems are very important and when there are relatively few traders, markets can lead to distorted outcomes. In laboratory experiments so-called 'combinatorial' auctions can be superior solutions to markets. This may be

important in the case of airport slots, whose sale may have to be coordinated, at least for the most important airports.

5. Auctions require the government to announce rules explicitly *ex ante*. This necessitates explicit statements of the type of services to be provided, and very often of the trade-offs between quality and costs in order to decide on bonus arrangements. Such explicitness should improve the rationality of public decisions.

There are three main costs of using auctions, which are mirror images of the benefits of negotiations:

1. Because rules must be announced well ahead, and cannot be changed, mistakes can be costly. For instance the sale of the right to broadcast television in Australia was marred by the fact that a firm took advantage of the rules in ways which delayed the end of the process. In New Zealand, the use of a 'second price' auction led to public outcry, as it gave the impression that the winners had underpaid for the resources that they obtained.
2. Auctions require that the auctioning body has good knowledge, when setting the rules, of the benefits of different options. It needs to be able either to put constraints on the use of the resources that will be acquired — you will have to run at least so many trains on the line during rush hour and outside of rush hour — or, even better but more difficult, to put a price on the value of some of the uses — a promise to run more trains during rush hour will be equivalent to an increase in the bid of so many euro per train. If it is not able to do this, it might prefer to ask potential participants for proposals that spell out the use of the resource and the reasons for which this use is preferable to other uses, and therefore proceed through negotiations.
3. When there are very few potential bidders, the optimal auction has a 'reserve price', that is a minimum price when a resource is being sold or a maximum price when governments are buying a service. The setting of this reserve price requires basically the same information than a successful negotiation.

#### 1.4. Administer

One should not forget that governments also directly regulate some activities, including activities that are important for network industries. For instance, even though road construction or maintenance can sometimes be done through the private sector, the job of integrating the network and planning it rests with civil servants. How much direct administration is to be done depends on the quality of the civil service in a specific country, quality which can vary over time.

## 2. Correcting market failures

### 2.1. Efficiency and markets

The great tradition of competitive analysis, starting with Adam Smith and cumulating in the crowning achievement of general equilibrium theory, has shown that competition, usually described as the presence of numerous small non-colluding firms, leads to efficient outcomes. On the other hand, we also know that there are circumstances in which competitive markets either will not function properly, or will lead to sub-optimal outcomes. As a consequence of these analyses, governments the world over have implemented policies that try to maintain competition where it is the efficient method of managing the allocation of resources, and policies that try to correct their outcomes when they are sub-optimal.

Most of economics studies the efficiency of markets and the methods by which governments should intervene. It is of course impossible to summarise adequately this huge body of knowledge in this report. We simply want to stress certain points that are especially relevant for network industries.

### 2.2. Promoting competition

#### 2.2.1. Antitrust

History shows that in most industries competition is not the natural state of affairs. For instance, the USA introduced antitrust law because many industries had become dominated by monopolies. The Swiss example shows that powerful cartels emerge when governments do not actively fight them. At the most elementary level, competition is unstable because firms that produce the same product have incentives to either merge or to come to pricing and production agreements in order to capture monopoly rents. The role of competition policy, in all industries, is to limit this phenome-



non. Network industries are no different from others in this respect, except for the fact that the public operator often continues operation after liberalisation, and may have incentives to maintain its market power by excluding competitors or by using his bargaining power to explicitly or implicitly collude with them.

### **2.2.2. Promoting entry**

If an industry with active competition is more efficient, it can be desirable for governments to promote entry when there are too few firms. In newly-liberalised network industries, regulators often favour entrants to achieve this goal. The experience of the UK telecommunications industry shows the pitfalls of this policy: the incentives of both the entrant and the firm in place are muted if it is known that the former will be protected. The regulator can be used as an instrument for implicit collusion.

## **2.3. Allocating property rights**

### **2.3.1. The problem**

Recent economic analysis has stressed the importance of the allocation of property rights for economic efficiency. Two problems of misallocation of property rights are common. First, some goods might have no owner, in this case there is a tendency for overuse, as agents do not take into account the cost that they impose on others. This is the case for environmental goods: there is overuse of rivers to carry away industrial or agricultural pollution because agents do not pay for this use of the resource. In the case of networks, there is more than socially-optimal congestion on roads at peak hours because commuters do not pay for the substantial delay that their presence on the road imposes on others.

### **2.3.2. Methods for allocating property rights**

Property rights can be allocated by all the methods discussed above in the section on contracts. Some experiments have recently been conducted on the use of auctions to assign property rights. This technique seems to be feasible in a wide variety of environments, and, if used intelligently, appears to be preferable to negotiations.

The most impressive use of auctions has been in the sale of the rights to use airwaves for telecommunications purposes in the USA, but they have also been used in New Zealand, and in Australia for allocating the right to broadcast television. The privatisation of rail transport in

Great Britain has also recently used an auction, as each section was sold to the highest bidder. There have been quite a few discussions on the auctioning of airport slots, but, as far as we know, no actual experiments.

## **2.4. Managing natural monopolies**

### **2.4.1. Technology**

#### *2.4.1.1. Traditional theory*

In some industries, it is technology itself, and not the incentives of producers, that endanger competition. Indeed, if large firms are much more efficient than small firms, i.e. if their costs are smaller than those of small firms, there will be space for only a small number of producers, or indeed in some cases for only one. We say that there are increasing returns to scale. It would be possible through regulation and laws to limit the maximal size, but the efficiency costs would be very high (although the evidence on increasing returns to scale in network industries is not very clear). The traditional analysis of network industries has stressed this phenomenon, and has argued that, because competition is unworkable, there should be only one firm. This firm cannot be left to maximise profits, as it would reduce output far below the efficient level, and it must be regulated in order to promote efficiency. The cut-off point at which it is considered preferable to enforce sub-optimal competition rather than regulate or administer depends in great part on the evaluation of the costs of government intervention, a topic that we discuss below.

#### *2.4.1.2. Scope of the argument*

Under which circumstance is it indeed necessary to intervene because of increasing returns to scale? Recent analysts have had a tendency to restrict the set of circumstances where they find the argument compelling, for different reasons, some which we find convincing, others that are less so.

#### *(1) Competition in some segments and not in others*

Many current reforms stem from the realisation that increasing returns to scale in parts of a network do not justify heavy government regulation of all the network. For instance, the British reforms of the electricity industry are based on the insight that transportation is a natural monopoly, but that generation is not. Similarly, the proposed reforms of the rail network in Europe are based on the assumption that one can separate the man-

agement of the tracks, a natural monopoly, from that of the rolling stock. It is difficult to know whether this recent tendency is the result of technological innovations that have allowed the introduction of competition where it was previously unfeasible, or whether distrust of government intervention has stimulated imaginations to find ways to limit it.

### (2) *Managing the interface*

The introduction of competition in segments of industries where other segments are managed by public or closely regulated industries creates another argument for government intervention, and this argument is very specific to network industries. As the theory of industrial organisation has shown, firms have incentives to transfer their monopoly power at one stage of production to other stages, either 'downstream' or 'upstream'. Competition at the interface must often be managed, and sometimes micro-managed. The consequence of this micro-management are clear in the very heavy involvement of regulators in the organisation of competition in some newly 'deregulated' industries. Whether this is a transitory phenomenon, which will disappear as the new entrants become stronger, is not clear at this point. There exist theoretical arguments on both sides of the issue. There is however a risk that competition will require heavy handed regulation for a rather long time.

### (3) *Contestable markets*

A number of authors have argued that even with strong increasing returns to scale, there need be no government intervention. Baumol, Willig and their co-authors <sup>(1)</sup> argue that even if increasing returns to scale imposed the presence of only one firm, the threat of entry could be enough to discipline it — this is the theory of contestable markets. Although this theory has been an important theoretical tool and has yielded useful theoretical insights, its practical relevance is nil. It is generally agreed today that even air transportation, which seemed the quintessential contestable industry, does not fall in this category.

## 2.4.2. *Contracts*

### 2.4.2.1. *Incomplete contracts and vertical integration*

Incompleteness of contracts arises when the parties cannot commit to all the actions that would be relevant to

come to a complete agreement. It is easier to start from an example of a complete contract. Placing a call from a public phone is a simple transaction where the operator promises to connect the customer with another party, while the customer promises to pay a certain amount. We say that the contract is complete because the two parties have no difficulty writing a contract that completely describes the transaction that they want to perform. When this condition does not hold, we say that contracts are incomplete.

Incompleteness of contracts arises in many transactions in the economy. It is for instance prevalent in labour contracts, and in this respect network industries are not different from other industries. On the other hand, they do create special problems for investment in network industries, for two reasons: these industries often use a large amount of specific capital, and the activities of the firms that are active in one network are, by definition, highly interdependent. Consider, for instance, an airline that wants to use an airport as a new hub. The airport needs to expand its facilities, including those used for air traffic control. On the one hand, the airline wants to ensure that the facilities are available before it redesigns its routes, on the other hand the airport will require guarantees on the revenues received from the airline. Any contract that tries to meet these two legitimate requirements will be highly complex, and will fail to some degree. For instance, the costs of the airport could depend in complex ways on the scheduling of planes, that are very hard to predict. Writing a contract that list all the relevant features of the schedule would be excessively cumbersome. There could also exist circumstances where it would be efficient for the airline to stop using this hub, and where the contract should allow it to do so, but these circumstances are very hard to predict and describe precisely enough *ex ante* that they can be written down in a contract. Incompleteness of contracts creates even more difficulties in industries where the technology changes very fast, such as the telecommunications industry.

The incompleteness of contracts typically leads to underinvestment in industries, such as network industries, where the capital is not easy to transfer from one use to the other, and firms often use vertical integration to protect specific investments. Regulators must intervene if they want to see both enough investment and competition in some segments in network industries.

<sup>(1)</sup> See, for example, Baumol, Panzar and Willig (1982).

#### *2.4.2.2. Managing the contracts to promote competition*

The judicial system is often not the appropriate forum to settle disputes stemming from the incompleteness of contracts. Indeed, the court will often ask itself the question ‘what is the agreement that the parties would have signed if they had predicted the new circumstances?’. This is very difficult to ascertain in the extremely complex environment of rapidly changing network industries. Given these difficulties, countries have found it reasonable to maintain a regulatory structure which aims, among other things, at maintaining rules that protect investment. For instance, a company that specialises in running trains on a network owned by another company will need to be protected against behaviour by the network operator that lowers the value of its investment. As usual, the fundamental question is to understand under which circumstances regulatory failures are less costly than the market failures that are being corrected.

### **2.5. Information management**

Asymmetry of information arises when one party is better informed than the other. The supplier of air travel is better informed about plane safety than the customer, and the seller of a new type of telephone is better informed about the reliability of the technology, or the probability that it will become obsolete in the near future. Asymmetry of information is pervasive in economic life, and has been studied extensively by economists in recent years. It is dealt with through contracts, reputation building, certification, warranties and so on. There does not seem to be any reason to believe that it is more pervasive in network industries than in other industries, although the asymmetry of information between regulators and firms create special problems, as discussed earlier.

### **2.6. Failure of the capital markets**

Most network industries require heavy investments and are therefore very dependent on capital markets. Some authors argue that private capital owners give too much weight to short-run profits, and that government intervention is required to take into account the long-term interests of society. There is actually no evidence that capital markets always have too short a time horizon (it has for instance been noticed that the American stock market reacts favourably to the announcement by a firm of increased investment in R & D) and the argument seems to be often used to soften the budget constraint of

government-controlled enterprises, whose current losses are excused because of some hypothetical future benefits. Furthermore, increased political power can also lead to ‘short termism’, as the next election becomes the horizon for decision making.

The fact that one should be careful about claims that public intervention is required in order to allow for an appropriate weight for the long-run future does not imply that there do not exist some real problems that private capital markets cannot handle. On average, network industries require longer-run financing than other industries, and long-run financial markets are more likely to be imperfect than short-run financial markets. In terms of modern theory, both long-run markets and long-run contracts are very incomplete. They are many contingencies on which no financial instruments can be based, and also many contingencies that agents find too costly to write in contracts. Governments may have some comparative advantages in correcting those failures, for instance through special taxation. Economic theory does not at this point provide precise guidelines on this point.

### **2.7. Network externalities**

Network externalities are discussed at length in Chapter I, Subsection 3.2, of this part of the report. At the general theoretical levels, they are simply special cases of a more general and well-understood phenomenon. If the decisions of an agent affect the welfare of other agents in ways that are not mediated through the price system, the market will not be efficient.

Externalities arise outside of network externalities. For instance, the creation of pollution by a firm generates externalities, and governments intervene, through taxation and regulation, to control these externalities. Similarly, there are positive externalities when a firm finances research and development, as part of the knowledge is made available freely to other economic agents. Once again, governments intervene, for instance by granting tax benefits. On the other hand, and practically by definition, externalities are more prevalent in network industries. Governments should intervene to manage them.

For externalities that are not specific to network industries, governments should use the same tools of control as in the rest of the economy. For instance, rail transport is often defended on the basis of its low impact of pollution. This should be reflected in taxes charged on

road transport, and the cost–benefit analyses of new investments should be based on those corrected prices.

On the other hand, there are network externalities in telephone service. When a consumer decides whether or not to connect to the phone network, he balances the utility that he will derive from this connection to the price he must pay. If his connection increases the welfare of other agents, because they can now reach him when they could not before, the cost benefit that he conducts does not include all the social benefits that it would generate. In some cases, he might decide not to purchase telephone service when it would be socially beneficial for him to do so. Consumption of telephone services would be too small. This can be one added justification for universal service.

Network externalities can go the other way. For instance, commuters choose the time at which they leave for work on the basis of the time it will take them to reach their office. As Vickrey (1955) has shown an extra car on the road at peak hour has very strong negative effects on the speed at which other vehicles travel. Because of the lack of an appropriate pricing scheme there is excess consumption of the good ‘travel at peak time’.

The intervention of governments to correct network externalities can be very complex, especially when competition has been introduced. Furthermore, it requires rules specific to the network in question. For instance, it is important for telecommunication networks to have common standards. They can be imposed by the regulator or chosen voluntarily by the firms that are active in the industry. One or the other of these solutions can be preferable. Negative externalities can usually be corrected through simpler methods, as the solution is usually to limit consumption, which can be done through the use of appropriate prices.

## 2.8. International considerations

### 2.8.1. Political risk

Another important issue which government must consider is national security. The analysis of the issue is quite complex, but the following example gives some flavour of the principles that are involved. There are two technologies for producing an output, technology A based on a risky foreign input and technology B, based on safe foreign inputs (one can think of electricity produced with imported petroleum or natural gas as tech-

nology A and nuclear power as technology B). If financial markets were working perfectly, private industry could handle the problem quite well. Firms would know that in case of disruption of foreign supply, the price of the output will increase. They would decide on their investment in one or the other technology on the basis of the expected revenue that they would generate, and it is easily shown that with perfect capital markets this would lead to an efficient allocation of resources. Two problems can arise.

1. If the private sector discounts the future too heavily, it will underinvest in the safe technology, whose benefits will only be available in the (uncertain) long-run future. As we have discussed in Section 2.6, this is probably not a good argument.
2. Industry might not trust the government to let prices rise in case of disruption. And indeed, even though letting the prices rise might be an efficient way to ration the limited supply, governments often find themselves under strong political pressure not to let ‘speculators profit from the situation’. Under these conditions, some government intervention might be justified in order to ensure that the proper mix of technologies is used, even if the government’s assessment of risks is similar to the market’s.

It must be recognised that this line of argument can be pushed too far: if national security is important, it must be acquired at the lowest possible cost. For instance, support for national airlines has often been justified on the basis of national security: the country needs to be able to control some means of air transport in case of international crisis. One can imagine other ways in which legal rights to use aeroplanes can be maintained in such a situation (for instance, ensuring that enough planes are leased to foreign airlines by nationals, and that the contract does specify some means by which they can be requisitioned in case of emergency).

### 2.8.2. Using market power

If the responsibilities of governments extended to the globe, they would try to promote competition. In a world of nation States, on the other hand, they realise that market power can be exercised by countries, and that they can have a role either in reinforcing the market power of their own industry, or in trying to limit the market power of other countries, who use it against

their own national interests. For instance, natural gas is traded in an oligopolistic world market. Consuming countries may have more bargaining power than private importers, and may rationally decide to use this bargaining power, by making the import of gas a State monopoly (as usual potential benefits have to be weighed against the cost of government intervention).

### **3. Income distribution**

#### **3.1. Markets and income distribution**

The third role of government is to redistribute income. This is a central issue in the politics of the regulation of network industries, albeit not in the writing of policy-makers and academics on this issue. For instance, the 'French system of public service' is often defended on the basis of equity issues.

##### **3.1.1. The problem**

The theory of income redistribution is simple. Some individuals are endowed at birth with greater earning ability than others, either because they are personal qualities highly valued by the job market, or because their family environment facilitates the acquisition of job skills, or because they inherit wealth that yields income. It is often argued that these endowments have no ethical values per se, and that society should try to level the playing field. (This is controversial: other authors would argue that there is an inalienable right to the fruits of one's own contribution to society, which should not be confiscated.)

##### **3.1.2. Taxation**

In the simplest framework, income redistribution should be conducted through lump sum transfers, that is transfers that are independent of any action of individuals. For instance, it would be announced that a tax would be levied on individuals of IQ greater than 125 to be redistributed to individuals of IQ less than 85. Such a scheme is, of course, not feasible, and rather than lump-sum transfers, governments use other instruments such as income tax or indirect taxes. It is often argued that taxes that do not distort the consumption choices of individuals are more efficient. For instance, income tax leaves consumers free to allocate their resources as they see fit. Similarly, a uniform VAT with only one rate modifies all consumption prices in the same proportion and should be non-distortionary.

Income taxes and one-rate VAT do however affect the efforts of consumers to increase their income. They can, for instance, induce agents to spend less time at work and more time in home improving activities. A general rule of economic theory states that in the presence of one distortion it can be optimal to introduce a second one, this is the theory of the 'second-best'. Translated in the present framework, it states that given that some distortions have been introduced in the choice between leisure and income, it could be optimal to introduce other distortions. For instance, in many countries, food is subject to a lower rate of VAT than other consumption goods. As food forms a large part of the budget of poor families, this is an efficient method for transferring income.

##### **3.1.3. Free provision of goods**

Governments also try to influence income distribution by subsidising or giving away some specific goods, and economic theory has shown that this can indeed be efficient. For instance, public access to education has an important redistributive aspect and governments provide it freely. Similarly, in all European countries access to medical care, at least for the more serious conditions, is essentially free (of course, citizens pay for it as taxpayers, but their payment is independent of their own consumption of medical services).

##### **3.1.4. Application to network industries**

Are network goods also a good vehicle for transferring income? First, it is clear that some network goods are non-negligible proportions of the budget of the poorest part of the population. They are therefore good candidates for favourable tax treatment. In some cases, pricing policies that would seem inefficient could actually be defended on these grounds. For instance, it is probable that local phone service forms a higher proportion of the telephone bill of households of modest means than of households of more substantial revenue. In this case, if we insist on budget balance within the telephone service, it could be appropriate to have local calls subsidised by long distance calls.

### **3.2. Merit goods**

The discussion up to this point has assumed that the aim of income redistribution was to limit the inequality of welfare in society. Under these circumstances, one simply tries to maximise the spending power of the poorest social groups. Some authors have argued that one should go further. There are some goods that are so



important for the integration of people in society, that there is a public interest in making sure that they are made available to the greatest possible proportion of the population at a low enough price that they consume a sufficient amount of the good. They are the so-called 'merit goods'. It is difficult to find examples of pure merit goods. In most cases, more standard economic arguments, such as externalities or contracting failures provide more compelling reasons for action.

### 3.3. Regional policy: income distribution among regions

Up to now, we have discussed income redistribution towards individuals. Its aim was to ensure that individuals had not too unequal levels of utility. Another tradition stresses the need to ensure that disparities between regions are not too sharp. Of course, income disparities between regions do not automatically translate into disparities between individuals, as migration compensates for its negative aspect. There must therefore be specific arguments for such a policy, usually aimed at favouring either rural districts or areas where established industries. Let us mention the most important.

1. There are negative externalities to large concentration of population; improving the standard of living in the poorest regions will stem out migration and lead to a general increase in welfare.
2. The feeling that the national territory should be evenly populated is sometimes taken as a basic value. If indeed the citizenship of a country feels that it is willing to devote resources to maintain, let us say, mountain agriculture, there is no fundamental reason why a policy aimed at maintaining it should not be considered legitimate. There is however scope for policy evaluation, as the costs should be explained as clearly as possible, and the fundamental political objective should be reached at lowest possible cost.
3. Finally, there is sometimes a feeling that every citizen, wherever s/he chooses to live has the right to be 'connected' to national life and to receive basic services from the State and the national community. This is not controversial for some services: for instance, it would be considered scandalous for some community not to be entitled to vote, because it is too far away and organising the election is too costly. Or, closer to

our preoccupation, police protection or access to the justice system are available in any location.

These considerations are often used to justify heavy investments, both by national and European authorities, in disadvantaged regions, and very often in networks: roads, telecommunications, airports. Two comments should be made. First, if we neglect the right to basic services, there exist other policies that can implement the objectives of regional policy. It has for instance been proposed that inhabitants of disadvantaged regions benefit from a lower tax rate. Second, investments in networks, especially networks that facilitate communications, can have the effect of lowering the population in remote regions. For instance, if the delivery of packages is improved, consumers will buy more clothing from catalogues, and this can lead to bankruptcy for local stores.

The third justification for the subsidising networks in disadvantaged regions is important should be viewed with some scepticism, as it is easy to use in self serving ways. More fundamental thinking about the issue is necessary, and this report deals with it in some more detail in its discussion of the universal service obligation.

## 4. The costs of government intervention

A number of authors have argued that even though monopoly power was inefficient, the remedy of government intervention is often worse. Traditional analysis of industries began by identifying 'market failures', and, as soon as one was identified, public intervention in order to correct it was deemed justified. Economists, and specially those from the 'Chicago school' have convincingly argued that government intervention was unlikely to be done perfectly well and one should weigh 'government failures' against market failures.

It should be stressed however that a list of supposed government failures is not sufficient to prove that government intervention is counterproductive. The policy literature seems to balance between two extreme positions. At one end, after a long list of market failures, the conclusion is drawn that energetic government intervention is needed. At the other end, one exposes all the reasons why government is inefficient, and then states the conclusion that the market should be left to do its

job. What is needed is a careful comparative assessment of market and government failures, which is often difficult to conduct. This comparison will yield different results in different circumstances, depending on the specific policies that are being implemented and the institutional framework.

#### **4.1. The efficiency of public institutions**

Some government failures are similar to market failures. For instance, the contract between a factory that produces electricity as a side product of its use of energy (co-generation) and a distributor is hampered by asymmetry of information. The optimal contract would have the factory supply electricity when its cost is less than the cost of alternative sources for the distributor, but neither of the parties knows the cost of the other. A government agency has no superior way to access either information, and there is no reason to believe that it could intervene effectively in this situation. (One should be careful to understand this statement precisely. The distributor could be trying to exercise his power as a monopsonist when negotiating this contract. There is a role in this case for antitrust policy. There is no role however for a policy specifically aimed at correcting the market failure.)

There is also evidence that regulated or public networks have more difficulties than most private firms to adapt to changing economic and technological conditions. The threat of competition is lacking and this absence helps employees, managers, suppliers and consumers resist changes that affect adversely, and they sometimes delay them far after their introduction would have been beneficial from the point of view of aggregate social welfare.

#### **4.2. Capture**

Some government failures are specific to government. The most often discussed is 'capture'. This concept, forcefully reintroduced in the modern discussion of industrial policy by Stigler (1971), stresses the fact that industries can influence policy makers to take decisions in their interests rather than in the social interest. This can be true both of potentially competitive industries (trucking, taxis) and of concentrated industries. In this view, it is important to draw clear line for government interven-

tion, as it will typically distort in favour of producers. The problem of capture is especially important in the decentralised environment of European industry, where the subsidiarity principle requires general guidelines decided at the Commission level to be implemented by governments and agencies of the Member States, who are often swayed by arguments of defence of national industries.

Capture has been very important in the history of network industries, in part because of the very high degree of government involvement. There is for instance clear evidence that the regulation of airlines in the USA up to the mid-1970s was rather inefficient, and more informal evidence that this was also true in most countries. Capture both by the companies, whether private or publicly owned, and by the employees of trade unions was at the root of the problem.

The solution to capture can be to reduce government intervention. In industries where a government will stay active as a referee if not as a direct actor, it is not clear that limiting government intervention to a few 'big' decisions, which is what is often meant by less influence, will reduce the cost of capture. For instance, the allocation of property rights can also be influenced by capture, as seems to be the allocation of slots at airports. Still, democratic control might be easier on a few large than on many day-to-day decisions.

#### **4.3. Income distribution from the poor to the rich**

In our discussion of income distribution, we have assumed that it went from the (relatively) rich to the (relatively) poor. In reality, income redistribution is subject to government failure in the same way than other policies. It is mediated through the political process, and also goes from the politically powerless to the politically powerful. Rich farmers do get subsidies.

Any instrument set in place for redistributing income will be captured at least in part by less-deserving groups. Furthermore, this income redistribution will often be done with highly inefficient instruments. For instance, work rules that yield minor increase in welfare to the employees are sometimes very costly in terms of global efficiency.

# Appendix: Public versus private firms — a survey of empirical tests of relative performance

## 1. Delineation

For a long time economists have debated on the issue of the respective merits of public and private firms. There is no clear consensus in the theoretical literature as to whether private monopolies are more efficient than public ones. This is partly why this question has motivated a large amount of empirical research since the 1950s. This effort has not faded in the recent years. On the contrary, the move towards more liberalisation and deregulation in various economic sectors has offered new opportunities of investigating this recurrent issue. The empirical literature remains inconclusive on the performance differential between public enterprises and privately owned regulated firms, for a given competitive and regulatory environment. However, efficiency is usually improved when competition is fostered while privatisation has ambiguous effects. Hence ownership has no impact per se, and cannot be gauged without taking into account other economic conditions. In other words, performance differential between public and private, when it exists, is due to a set of institutional constraints and incentives. This conclusion is widely recognised by recent theoretical approaches <sup>(1)</sup>.

There are several strands of literature that discuss theoretical arguments on the likely effects of ownership on productive efficiency. A first view, developed by Leibenstein (1966) in a well known article, predicts no efficiency differential on the basis of type of ownership. Public and private firms should be equally X-inefficient, although the inefficiency may not be of the same nature as is observed from anecdotal evidence. On the contrary, the property rights view pleads for an efficiency advantage of private firms over public firms. Alchian (1965) identified the crucial difference between

public and private firms as the high cost which public ownership imposes on the transfer of property rights pertaining to the firm. This cost is due to the lack of market for public shares and to transaction costs. In addition, public choice theories provide several arguments showing how political optimisation can conflict with cost efficiency. In particular, politicians and bureaucrats who are in charge of monitoring the performance of public firms, may choose production levels that maximise their own objectives, which can hardly produce a social optimum. These different arguments on the effect of ownership on productive efficiency show that it is a matter of empirical research.

This presentation of the theoretical literature on the effect of ownership on efficiency is rather simplistic. As a matter of fact, economic reality is complex and requires several lighting-up angles. There are no clear-cut arguments indeed as the various approaches are replies of one another. However, professional economists should agree that the ‘principal–agent’ relationship between the manager and the owner is at the core of the question of assessing the performance of a firm and that the ‘principal–agent’ problem is just as great in private firms as in public enterprises. Then the effect of ownership on efficiency cannot be measured independently of the institutional constraints and the system of incentives.

Several remarks or examples can support this assertion. First, within a setting of uncertainty and asymmetric information, competition can provide the pressures for inducing managers to operate as closely as possible to their production frontier and the relevant information for better monitoring their activities. In this context, measures of efficiency could be the basis for implementing reward schemes designed for inducing efficiency. As a market system creates interdependence between managers’ utilities, it also provides an incentive scheme. It is often observed that, in the experiences of privatisation as reported by Kay and Thompson

<sup>(1)</sup> See Laffont and Tirole, 1993, Chapter 17, p. 637; Vickers and Yarrow (1991), p. 116.



(1986), the greater the competitive pressure exerted on managers in an industry, the more difficult it is to secure the support of public sector managers for ownership transfer. To gain support for a privatisation programme, one then could need to reduce competitive pressure in the post-privatisation market. Second, the ownership effect depends on details of the regulatory environment. The obvious example is that the deregulation of previously regulated private firms often favours efficiency. Moreover, when comparing a public and a private regulated firm, as stressed by Laffont and Tirole (1993), one has to contrast the cost of public ownership due to the need to allocate resources away from profit-maximisation purposes (as a public firm should have welfare-enhancing objectives) with the cost of private ownership arising from the multiplicity of principals (shareholders and regulators). According to different studies, privatisation may or may not improve efficiency.

The empirical tests on the effect of ownership on efficiency have then to take into account the close relation between ownership, competition, regulation and various environmental constraints. However there are practical difficulties to do so. For sectors dominated by public monopolies like postal services or railways, it is obviously difficult to perform a statistical analysis of the effect of ownership. One may study the temporal pattern of efficiency of these firms, or the effect of privatisation on their cost efficiency, or one may perform an international comparison in order to assess the effect of a larger and tighter autonomy of public services in different countries. Empirical studies on the performance differential between public and private firms are restricted to activities where the two types of firms are present. This is often the case for sectors like insurance, banking, refuse collection, water distribution, electricity distribution, etc.

Comparisons over time, across space, across economic sectors, among different types of ownership and regulation usually provide the materials of performance studies. Over the last 20 years, many studies have compared private and public firms on the basis of average costs, which often was to the advantage of private firms <sup>(1)</sup>. However, the criterion of average costs is not able to measure tech-

nical and allocative efficiency as soon as prices do not reflect the social values of goods and services. The next part discusses what is meant by performance and proposes to use a measure of productive efficiency derived from the so-called production frontier models. Afterwards a survey of empirical studies is presented.

## **2. Defining and measuring economic performance**

---

Comparing profit levels is the most classical way of assessing the efficiency of private firms. The profit is the unique objective of shareholders or owners. In the case of public firms, it is often recognised that the performance must be gauged with respect to the achievement of multiple objectives. (See Pestieau and Tulkens, 1993, on this point.) However obtaining a unique and global measure of performance is in general too ambitious because the manifold objectives may not be compatible with one another and because usually data are not available to undertake such a project.

It has long been recognised that technical efficiency is the relevant criterion for comparing public and private firms. First, it is the sole objective which does not prevent the achievement of other objectives. Indeed, being technically inefficient cannot be justified on the basis of other objectives. Allocating too many resources to a production process for social or environmental reasons (for instance) does not necessarily imply that the allocation is technically inefficient. Then a measure of technical efficiency should not be polluted by the presence of multiple objectives as in the case of public firms. This is also true for private firms. While they should be both technically and allocatively efficient in a competitive world, they can be (first-best) allocatively inefficient due to informational asymmetries between owners and managers. Second, statistical data on outputs and inputs are often available at the firm level for all types of enterprises, which allows for the evaluation of the so-called production frontier, i.e., the interior of the production set.

Farrell (1957) was the first to propose the idea of measuring productive efficiency of individual decision units and to distinguish two mutually exclusive and exhaustive sources of productive inefficiency: technical and allocative inefficiency. In microeconomic theory a production function is defined in terms of the maximum output that can be produced from a specified set of inputs, given the existing technology available to the firms involved, or vice-versa, in terms of the minimum

<sup>(1)</sup> Gathon and Pestieau (1992) survey all the studies bearing on the sector of refuse collection since 1965. The contrast between the results obtained from the studies using the average cost as the measure of efficiency and the more recent studies based on the measure of technical efficiency is particularly striking.

of resources employed for producing a certain level of output. Graph 4 depicts the situation in which firms use two inputs of production, K and L, to produce their output. Each firm is represented by a full circle. The production function is represented by the isoquant ( $Q$ ), i.e., the boundary set of input combinations which can be used together to produce a given level of output. In other terms the isoquant defines the locus associated with the most efficient use of the inputs to produce the output involved. The deviation from the isoquant provides a measure of technical efficiency. If the firm being analysed at R on Graph 4 was on the isoquant at T, it would be 100 % technically efficient. Then the ratio  $OT/OR$  is the measure of technical efficiency. Technical inefficiency thus arises from excessive use of inputs and hence the actual bundle of inputs used lies within the best practice isoquant.

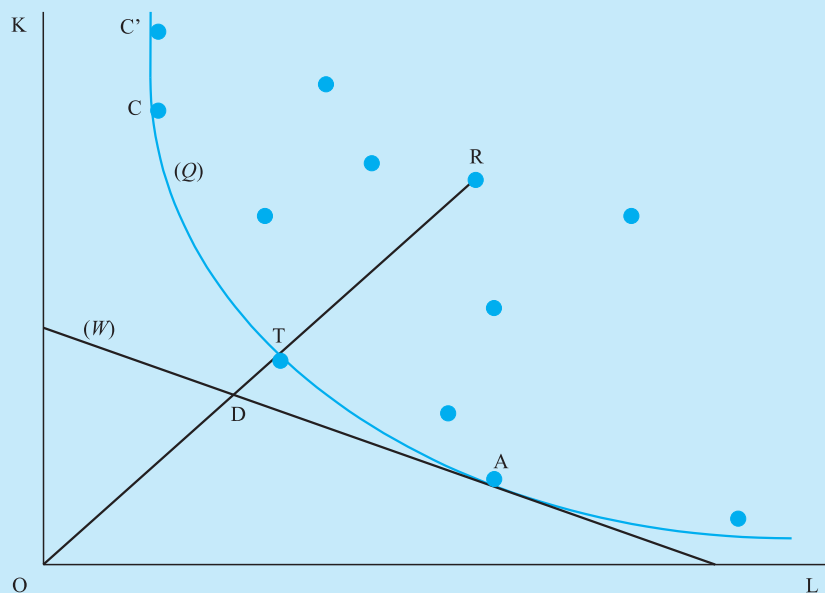
Allocative inefficiency arises because the ratio of inputs does not reflect the ratio of relative input prices and hence total cost could be reduced by employing a different mix of inputs. Given that the unit is technically efficient (for instance, point T on Graph 4), allocative

inefficiency arises because the unit is producing at the wrong point (point T) on the frontier given the relative factor prices that it faces. In other terms, allocative efficiency captures the inefficiency arising solely from the wrong choice of technically efficient combinations given input prices. Given that the line ( $W$ ) on Graph 4 represents the cost level associated with the different combinations of input quantities for a given level of the factor price ratio, the cost levels in A and in D are equal. So when a firm operates in T instead of A (which are both technically efficient allocations), a measure of allocative efficiency is given by the ratio  $OD/OT$ .

Thus given the available production technology ( $Q$ ) and input prices ( $W$ ), efficient (minimum cost) operation occurs at A. With respect to the unit to be analysed, operating at point R, the distance  $OD/OR$  denotes the measure of overall efficiency. This measure can be disaggregated into two components, namely technical efficiency and allocative efficiency. The measure of overall efficiency is the product of the two other measures as:

$$OD/OR = (OT/OR) \times (OD/OT).$$

Graph 4: Overall, allocative and technical measures of productive efficiency



Note the radial nature of these measures. This imposes some limitations. Indeed points C and C' on Graph 4 are both technically efficient with respect to the radial measure, but not as efficient in a Pareto sense. Indeed C' involves a quantity of one input (K) strictly greater than C for producing the same output level. Although some non-radial measures have been proposed in the literature, they are less convenient since they do not imply the decomposition presented above.

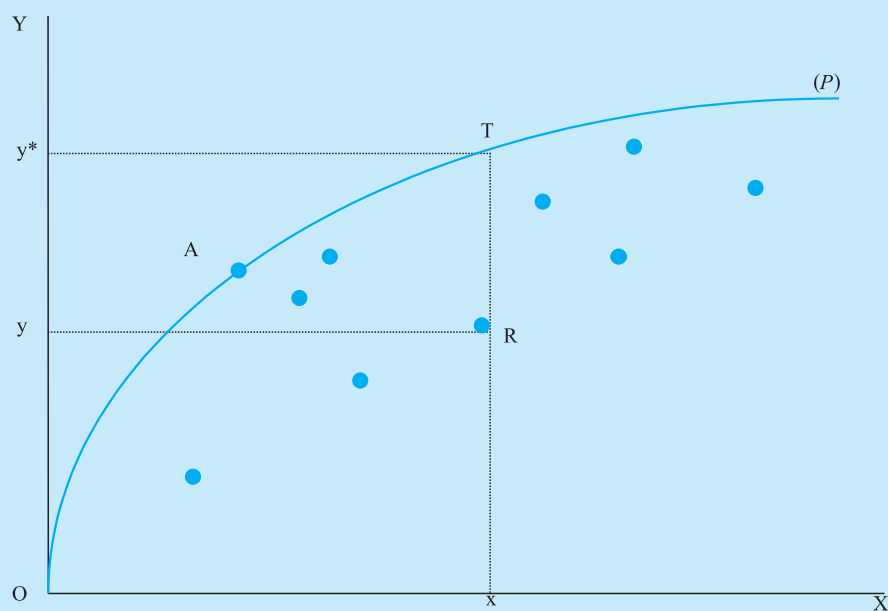
The goal of performance studies is then to measure technical or productive efficiency. Given the above definitions, whatever the ultimate objective of the firm or the individual decision unit, it has to be technically efficient. Moreover, being technically efficient is compatible with the other objectives assigned to the firm. Comparing firms which may have different objectives on the basis of a measure of technical efficiency is then quite convenient. At least, it is moving in the right direction, particularly when one comes to the problem of comparing the performance of private firms with that of public firms which may have to satisfy social objectives for instance. However, obtaining a correct mea-

sure of technical efficiency is not so clear in practice. The discussion on the definitions of efficiency also shows that the position of a firm in the input space may be the result of a complex combination of technical and allocative inefficiencies. It is then a challenging empirical task to identify purely technical inefficiencies, i.e. not blurred by allocative inefficiencies. This explains in part why the empirical literature on the measure of efficiency is so abundant and why many measurement methods have been investigated.

### 3. The production frontier approach

All methods proposed so far have in common to start with the characterisation or the estimation of a production frontier associated with the case to be analysed. This concept of production frontier is depicted in Graph 5 involving the original input and output values. The horizontal axis represents the inputs, X, associated with producing the output, Y. From available statistical data that report on the outputs achieved and the inputs used by an individual decision unit (represented by each full circle on Graph 5), the production frontier must be

Graph 5: Technical measure of efficiency in input–output space



inferred or constructed, before any efficiency computation can be made. The observed input–output values are in general below the production frontier ( $P$ ), given that firms do not attain the maximum outputs possible for the inputs involved, given the technology available. Efficient firms are those operating on the production frontier such as the firms being located at point A. A measure of technical efficiency of the firm which produces output,  $y$ , with inputs,  $x$ , denoted by point R, is given by the ratio  $y/y^*$ , where  $y^*$  is the frontier output associated with the level of inputs,  $x$  (see point T).

To estimate production frontiers, several methods have been proposed, usually classified as parametric or non-parametric. The main difference between these two categories of methods is the technique used to describe the frontier. The non-parametric methodology, solves a system of linear equations for each sample observation, that satisfies some desirable properties of production possibility sets, such as convexity, free disposal, etc. Use of non-parametric methodologies mostly prevails in the management science and operation research literature. (See Ali, Lerne and Seiford, 1995.)

Parametric methods found in the econometric literature are based on the statistical estimation of a parametric function that links the output to the inputs through constant parameters, which are evaluated in the process of estimation. They provide the statistical function that best envelops the data from above, the so-called stochastic production frontier. The measure of technical efficiency is then obtained by looking at the ratio between the observed output and the output predicted by the estimated function, which corresponds to the maximum output possible given the data available. The quality of the measure clearly depends on the chosen functional form, but also on the stochastic assumptions required to perform the statistical analysis (<sup>1</sup>).

The stochastic production function is defined by

$$y_i = f(x_i, z_i; \beta) + \varepsilon_i - u_i \quad i = 1, 2, \dots, N,$$

where  $y_i$  represents the production level for the  $i$ -th sampled firm,  $f(x_i, z_i; \beta)$  is a suitable function of the

vector,  $x_i$ , of inputs for the  $i$ -th firm, a vector,  $z_i$ , of environmental variables, and a vector,  $\beta$ , of unknown parameters. Several remarks can be made on this model.

- The stochastic part has two components. The first one,  $\varepsilon_i$ , is the usual disturbance introduced in regression models and represents all types of omitted or unobservable variables that have unbounded effects on output (such as weather uncertainty, measurement errors on output, etc.) in production frontier models. The second component,  $u_i$ , is supposed to be non-negative valued random variable ( $u_i \geq 0$ ,  $i = 1, 2, \dots, N$ ) expressing unobservable factors, such as ability, effort, or technical practices of managers, that have bounded effects on outputs. As these factors do not reach their optimal levels, they give rise to inefficiency.

- Technical efficiency of an individual firm is defined in terms of the ratio of the observed output to the corresponding frontier output, given the levels of inputs used by that firm. Usually the absolute measure is obtained as  $\exp(-u_i)$ , and the

relative measure as  $\exp\left(\frac{\max_{j=1}^N u_j - u_i}{\max_{j=1}^N u_j}\right)$ ,

which is lower than one.

- The discrepancy between the best-practice frontier and the observed level of output can be caused by many factors. When exogenous variables like  $z_i$  (bearing on geographical, institutional, regulatory factors, etc.) are available, one is able to disentangle their effect from what is solely under the responsibility of managers. Then managerial inefficiencies can be separated from inefficiencies due to other factors.

Various extensions of this preceding canonical model have been considered and various methods have been applied to estimate the model parameters. The choice among them is mostly driven by the type of data available. One important technical difficulty of this type of model is that input quantities can be correlated with the error term representing managerial inefficiencies. Indeed, at the equilibrium, the level of factor prices, which determine the level of input quantities, may also affect the effort of managers in achieving efficiency.

(<sup>1</sup>) This is why, more recently, non-parametric statistical methods have been developed in order to avoid to specify a particular parametric form for the production function and to let the data decide. These methods emphasise the statistical nature of the data. See, among others, Cazals, de Rycke, Florens and Rouzaud (1997).

There are different solutions to deal with this question <sup>(1)</sup>. They depend on the type of data available, cross-sections of several productive units, time series of observations of the same unit, or panel data, i.e. series of cross-sections <sup>(2)</sup>.

Analysts are still debating on the respective advantages and inconveniences of parametric and non-parametric methods to estimate the production frontier. The main difference between these methods is that parametric methods recognise the stochastic nature of data and allow for random components, while non-parametric methods are completely deterministic and are very sensitive to outliers. However parametric methods usually require to construct a one-dimensional index of output, which implicitly involves a particular choice of aggregation technique among outputs when one deals with multi-output firms. Moreover, they also require the choice of a parametric form to approximate the true production function, which can introduce specification bias <sup>(3)</sup>. It is not the case with non-parametric methods since the basic idea of these methods is, in some sense, to let the data choose the aggregation principle and the shape of the production frontier. One understands then why the present line of research in this area of production analysis is to combine the advantages of non-parametric and parametric methods.

Obviously both methodologies can be applied to the same data set. When applied to the same case, either they consolidate between themselves when they provide the same type of conclusion, or they allow to discover new insights when they provide different answers.

#### **4. Survey of empirical studies**

Although the empirical literature bearing on the measure of technical efficiency in various economic sectors is abundant, it is not possible to cover all network industries of the main EU countries. Here the review of the empirical studies attempts to cover these industries and related sectors where the service is provided by some forms of public enterprises like cooperative, self-managed firms, non-profit organisations, local authorities, etc.). Comparisons between countries where a specific market (like airlines) has been strongly deregulated

and countries where public companies remain powerful are also included, as they shed light on the relation between efficiency and ownership. The most important task is to understand the causes of performance differentials between private and public firms, but quite often the sources of slacks are missing in the empirical studies.

Table 17 is reproduced, with some slight amendments, from Pestieau and Tulkens (1993). All cases can be encountered. Two studies show that American airlines are more efficient than French airlines. Another one indicates that public electric utilities are more efficient than private ones, but ownership does not seem to explain the differences. In refuse collection, tendered services have higher efficiency ratings than non-tendered ones. Insurance mutuals are doing better than public and private organisations.

Table 18 bears only on European railways. Pestieau and Tulkens (1993), who have produced this table, are able to separate the effect of regulation on the efficiency measures. The most striking result is that the Finnish company, the less autonomous from the central government, is the most efficient in terms of managerial efficiency. In other terms, if it was fully autonomous (i.e. private), then it would be the most efficient. Note that the Swiss firm, which is 100 per cent autonomous, does not perform very well.

However, one has to be very cautious in using these results. Cowie and Riddington (1996) compare different studies bearing on European railways (see Table 19). It is easy to see that all studies do not agree on the ranking of the different companies. They argue that, while the quality of data is poor for all these studies, one should explain these differences by problems of specification. In particular, the authors attribute to some methods of estimating productive efficiency and/or to model specification the fact that, in their own study on efficiency of railways, the Danish railways are more efficient than the British system while the former is more centralised than the latter.

A theoretical example, proposed below in a note, argues that the production frontier approach for estimating measures of efficiency can be justified, but that it is not able to provide unbiased efficiency measures because it does not take into account the regulatory constraints in a proper way. This note should motivate further research.

<sup>(1)</sup> See, among others, Ivaldi, Monier-Dilhan and Simioni, 1995.

<sup>(2)</sup> For technical details, one may refer to Schmidt, 1986, Bauer, 1990, and Färe, Grosskopf and Lovell, 1994.

<sup>(3)</sup> See, on this point, Florens, Ivaldi and Larribeau, 1996.

## 5. Conclusion

This survey explains that performance cannot be directly related to ownership. It means that competition could increase performance, whatever the type of ownership. Improvement of efficiency in public firms must be achieved through the system of incentives and institutional constraints.

This survey also suggests that, while the usual methods for measuring efficiency are widely applied, they do not take into account the environmental and institutional constraints impinging on the activity of firms in a proper way. It advocates the use of structural econometric models.

## Note

We present below a simple model which aims at supporting the conjecture that technical efficiency cannot be measured independently of the institutional or regulatory constraints. The organisation of urban transport in France serves as the background of the model.

Consider an urban transport network. The local authority (LA, hereafter) is the principal while the network operator (NO), which can be a public or private firm, is the agent. LA sets the prices  $p$  of public transportation and chooses the clauses of the management contract with NO. Assume that there are two types of contract: Cost-plus contract (CPC) and fixed-price contract

Table 17

### Productive efficiency — comparative studies of public and private firms

Sector and authors	Number of units	Type and period of data	Number of out-puts and inputs	Method	Mean efficiency degrees	Remarks and other findings
Airlines						
Good, Roeller & Sickles (1991a)	9 airlines	Panel annual 1976–86	1 output 4 inputs	Parametric	—	— US airlines are 10–15 % more efficient than French airlines.
Good, Roeller & Sickles (1991b)	16 airlines	1976–86	1 output 4 inputs	Parametric and non-parametric		— US carriers are 15–20 % more efficient than European carriers. — Over the period, European technical efficiency declined.
Barla & Perelman (1989)	26 airline companies	Panel annual 1976–86	1 output 2 inputs	Parametric	About 80 %	— Main purpose is to compare efficiency performance of European and American deregulated companies. Public companies are performing better during crisis periods while deregulated ones are performing better in favourable economic conditions.
Manzini (1990)	50 airline carriers	Cross-section 1987	1 output 2 inputs 4 output attributes	Parametric	73.1 %	— Main purpose is to estimate a production function in order to get technical information on the production process and to assess the influence of property rights and regulation on technical efficiency. American carriers are more efficient.
Banking						
Tulkens (1993)	773 branches of a public bank and 911 branches of a private bank	Monthly data 1987	7 outputs 1 input	Non-parametric	About 97 %	— Proportion of branches on the frontier: 75 %
		1986	7 outputs 1 input	Non-parametric	About 95 %	— Proportion of branches on the frontier: 58 %
Education						
Rhodes & South- whick (1988)	64 public and 57 private universities in the US	Panel annual 1971, 1974, 1981	5 outputs 5 inputs	Non-parametric	About 88 % a year	— Private universities have slightly higher efficiency scores, for every year considered.



<b>Electric utilities</b>						
Färe, Grosskopf & Logan (1985)	30 public and 123 private utility plants in the US	Cross-section 1970	1 output 3 inputs	Non-parametric	—	— Public plants have better ratings in terms of technical efficiency measures than private ones. Congestion is more a problem for public than for private utilities.
Hjalmarsson & Veiderpass (1991)	289 Swedish retail electricity distributors	Annual data 1970–86	4 outputs 4 inputs	Non-parametric	—	— Ownership or economic organisation does not seem to be related to productivity change in any significant way. — The municipality and the state owned companies display the highest efficiency values during most years.
<b>Hospitals</b>						
Grosskopf & Vladamis (1987)	22 urban public and 60 private hospitals in California	Cross-section 1982	4 outputs 4 inputs	Non-parametric	About 97 %	— Sample partitioned into public and private hospitals. — Comparison of results of each hospital with the whole sample frontier and the separate ownership frontier. — Best-practice frontiers differ, public hospitals using less resources. — Public hospitals are more efficient in both samples.
<b>Insurance</b>						
Fecher, Kessler, Perelman & Pestieau (1993)	243 French non-life and 84 life companies	Panel 1984–89	3 outputs 2 inputs	Non-parametric	At most 50 %	— Public companies outperform private companies and mutuals.
Delhaussse, Fecher, Perelman & Pestieau (1991)	243 French and 191 Belgian non-life companies	Panel 1984–89	1 to 3 outputs 2 inputs	Parametric and non-parametric	At most 50 %	— Non-profit companies are more efficient than for-profit ones. French companies are more efficient than Belgian ones.
<b>Railways</b>						
Oum & Yu (1991)	21 railway companies	Annual data 1978–88	1 output 5 inputs	Non-parametric		— Managerial freedom (autonomy and government agency variables) has significant effect on efficiency. — The dummy variable for quasi-public firms (partial private ownership) is not statistically significant.
Filippini & Maggi	57 railways under mixed ownership	Annual data 1985–88	1 output 3 inputs 2 networks characteristics	Parametric	1 each year	— Limited evidence has been found for a relationship between the share of state in capital and cost efficiency. — Positive correlation appears between cost efficiency and the importance of the canton's participation in the deficit of firms.
<b>Refuse collection</b>						
Cubbin, Domberger & Meadowcroft (1987)	Refuse collection in 317 local authorities	Cross-section 1984–85	10 outputs 2 inputs	Non-parametric	81 %	— Tendered services have higher efficiency scores than non-tendered ones.
Burgat & Jeanrenaud (1990)	Refuse collection in 98 municipalities in Switzerland	Cross-section 1989	2 outputs 2 inputs	Parametric and non-parametric	—	— Technical gains can be obtained by contracting out the service to a private collector.
Distexhe (1993)	Refuse collection in 176 Belgian municipalities	Cross-section 1989	2 outputs 2 inputs	Parametric	75 % (public) 72 % (private)	— Productive efficiency increases with biddings.

Table 18

Efficiency of European railways

Railways	Country	Autonomy (%)	Railways	Managerial efficiency	Efficiency measures 1986–88		Railways	Gross efficiency
					Railways	Regulatory efficiency		
BLS	Switzerland	100.0	VR	1.000	BLS	1.000	NS	0.897
BR	UK	76.3	TCDD	0.987	SJ	0.955	TCDD	0.888
CFF	Switzerland	66.0	NS	0.964	BR	0.946	CFF	0.883
CFL	Luxembourg	63.5	CIE	0.963	NS	0.930	BR	0.877
CH	Greece	47.3	CFF	0.962	SNCF	0.929	SNCF	0.875
CIE	Ireland	58.3	SNCF	0.942	FS	0.921	BLS	0.873
CP	Portugal	64.0	RENFE	0.939	CFF	0.918	CIE	0.862
DB	Germany	61.0	OBB	0.939	SNCB	0.914	CP	0.856
DSB	Denmark	41.5	CP	0.938	CP	0.912	SJ	0.835
FS	Italy	67.0	BR	0.927	CFL	0.0911	FS	0.830
NS	Netherlands	70.3	FS	0.901	DB	0.903	VR	0.828
NSB	Norway	45.3	SNCB	0.890	TCDD	0.900	RENFE	0.822
OBB	Austria	41.8	DB	0.888	CIE	0.895	SNCB	0.813
RENFE	Spain	52.3	SJ	0.875	RENFE	0.875	DB	0.802
SJ	Sweden	80.0	BLS	0.873	CH	0.857	OBB	0.785
SNCB	Belgium	64.5	CH	0.867	NSB	0.850	CH	0.743
SNCF	France	69.8	DSB	0.830	OBB	0.836	CFL	0.731
TCDD	Turkey	60.0	NSB	0.827	DSB	0.834	NSB	0.702
VR	Finland	40.0	CFL	0.803	VR	0.828	DSB	0.693

Source: Pestieau and Tulkens (1993).

Table 19

Comparison of efficiency measures for railways

Country	Efficiency study with period of data				
	PP, 1982–84	GPs, 1986–88	GP1, 1961–88	GP2, 1961–88	OY, 1989
Austria	49.6	88.8	61.0	63.3	85.0
Belgium	37.6	86.7	88.5	87.0	73.0
Britain	50.4	98.7	71.9	73.5	100.0
Denmark	71.9	83.0	82.6	98.0	82.0
Finland	55.6	80.3	91.7	100.0	96.0
France	57.2	83.0	100.0	99.0	99.0
Germany (W)	48.3	93.9	82.6	98.0	91.0
Ireland	70.2	94.2	71.9	73.0	100.0
Netherlands	98.7	90.1	94.3	99.0	100.0
Norway	65.0	89.0	89.3	91.7	94.0
Spain	50.0	87.5	75.8	77.5	100.0
Sweden	59.7	87.3	75.2	97.1	100.0
Switzerland	56.2	96.4	81.3	75.2	89.0
<b>Average</b>	<b>59.3</b>	<b>89.2</b>	<b>82.0</b>	<b>87.1</b>	<b>93.0</b>

Source: Cowie and Riddington (1996).



(FPC). With CPC, LA recovers total sales, reimburses *ex post* costs and provides a monetary transfer  $t_o$  to balance the budget. Under a FPC, LA obtains all revenues and reimburses costs *ex ante*, i.e., provides a transfer  $t_o$ , but OP bears the losses or keeps the profits. Clearly, a FPC produces higher incentives than a CPC.

Assume that the LA cannot observe the quantity of labour. Specifically, if  $L$  denotes the observed labour force, the actual quantity of labour is:

$$(a.1) \quad \tilde{L} = L \exp(a - \theta),$$

where  $\theta$  is a random variable representing the efficiency of labour (a private information of OP) and  $a$  is the effort of managers to improve the efficiency of the firm through the management of labour. The technological conditions of production are described by the function:

$$(a.2) \quad y = f(K, \tilde{L}; \beta),$$

where  $y$  is the production level,  $K$  is the capital factor, and  $\beta$  is a parameter defining the technological conditions. The total cost is:

$$(a.3) \quad C = rK + wL.$$

Let  $\rho$  be a dichotomous variable, taking the values 0 or 1 for a CPC or a FPC respectively. Given that the regulator sets prices, the production level is obtained through the inverse demand function. Hence, the OP's programme is defined as follows:

$$(a.4) \quad \begin{aligned} \text{Max}_{a, K, L} \quad & t_o + \rho[py - rK - wL] - \psi(a) \quad \text{s.t.} \\ & y = f(K, L \exp(a - \theta); \beta), \end{aligned}$$

where  $\psi(a)$  is the cost of effort (assumed to be a convex function). The solution of this programme is given by the functions:

$$(a.5) \quad \begin{cases} a^* = \begin{cases} a(p, y, w, r, \rho; \beta) & \text{if } \rho = 1 \\ 0 & \text{if } \rho = 0 \end{cases} \\ K^* = K(p, y, w, r, \rho; \beta) \\ L^* = L(p, y, w, r, \rho; \beta) \end{cases}$$

Suppose now that we want to estimate the technology using a production function approach. For this purpose, one needs to specify the production function. For simplicity, one considers a Cobb-Douglas specification. Given the above model, the equation to be estimated is:

$$(a.6) \quad \ln y = A + \beta_1 \ln K^* + \beta_2 \ln L^* + \beta_3(\theta - a^*).$$

One may add an other term to account for probable measurement errors on the production level. However, the striking fact here is the similarity of equation (a.6) with the usual model estimated in the production frontier approach and given in the main text under paragraph 3.4. Indeed, the last term of equation (a.6) looks like the usual decomposition of error terms followed by the production frontier approach. In other terms, with an asymmetric information model, we have been able to explain why this decomposition is meaningful. Moreover, the above discussion also explains why error terms are structurally correlated with the input variables in a production function. So the econometrician has to take care of a serious problem of endogeneity when estimating a production function, a very classical problem which is worthwhile to keep in mind. In addition, equation (a.6) justifies the usual interpretation of disturbances in production frontier models in terms of unobservable inputs, like the quality of the management or the quality of some inputs. On the other side, the above example indicates why it is useful to develop a structural analysis to well identify productive efficiency: the error term must depend on the technological and regulatory conditions in a very specific way.

By the preceding discussion, we argue and show that one must be very cautious in using econometric results on efficiency measurement, as long as regulatory constraints are not taken into account in a proper way.

# III. Competition policy versus regulation

## Introduction

The opening of network industries to competition has paved the road for a possible replacement of the traditional regulatory paradigm by competition policy. While this substitution has yet failed to materialise, there are signs that it may be on its way. Regulatory decisions are often referred to and scrutinised by antitrust authorities and courts in Anglo-Saxon countries. In New Zealand, regulatory agencies in most network industries have been abrogated altogether, leaving to sections 36 (on the abuse of market power) and 27 (on illicit vertical restraints) the control of anti-competitive behaviour. And many telecommunications experts predict that the regulatory environment as we know it is bound to disappear in the wake of generalised competition. But there is widespread agreement that regulation will be needed in the short and medium term during the transition toward full-fledged competition. Furthermore, the conceptual framework for long run competition in those industries is still to be elaborated, as we discuss in Chapter IV, Subsection 4.

These developments raise an important institutional design question: which of the regulatory agency and the antitrust authority is best placed to handle the complex environments of liberalised network industries? Are regulation and antitrust substitutes (which calls for some antitrust exemption of regulatory decisions or conversely for the dismantlement of regulatory agencies) or complements? Before embarking on an analysis of these questions, we ought to issue two caveats. First, there is no such thing as ‘regulation’ on the one side and ‘antitrust’ on the other side. Rather, ‘industry oversight’, as we will call it broadly, has many facets. There are therefore many forms of regulation and many forms of antitrust enforcement, as well as alternative but related methods of managing the industry, such as compulsory arbitration. Furthermore, although we will need to draw as sharp a contrast between regulation and antitrust as possible for the purpose of the analysis, we will also have to acknowledge a number of points of convergence between

them. Second, we are not aware of a full-fledged analysis of this particular institutional design problem. Developing such an analysis lies beyond the scope of this report. At most can we bring recent developments in the economics of information and incentives to bear on the issue. We will thereby clarify the debate but will stop short of bringing a definitive answer to the question.

We proceed as follows. Starting from first principles we recall the *raison d’être* of industry oversight. We then attempt to characterise ‘regulation’ and ‘antitrust’ along a number of dimensions: procedures and control rights, timing of oversight, information intensiveness and continued relationship, and independence vis-à-vis the political environment. Last, we ask how these institutional features contribute to the fulfilment of the primitive objectives we started from.

## 1. The objectives of industry oversight

Let us recall the three broad objectives behind government intervention in industries (see Chapter II for a broader discussion). Taking service offerings as given, the first goal is to promote productive (cost and investment) efficiency without creating excessive rents. It is tempting to list efficiency and rent extraction as separate objectives. There is however a fundamental trade-off between them <sup>(1)</sup>. Schemes that let the firm share its revenue or cost with consumers or the government reduce the firm’s incentive to operate efficiently; conversely, the provision of ‘high-powered incentives’, namely of incentives that make the firms accountable for most of their cost and revenue, generate large variations in profits, leading to unfeasibility (bankruptcy or refusal to operate) or conversely to large and distortionary (as well as politically unsustainable) rents <sup>(2)</sup>.

<sup>(1)</sup> See Laffont and Tirole (1986, 1993).

<sup>(2)</sup> The distortion stems from the fact that these rents must be paid either through high, consumption-reducing rates or through government subsidies that increase the deadweight burden imposed by taxation on the economy. The incentives-rent extraction trade-off has sometimes been misappreciated, in particular by those advocating miraculous new schemes.

The second objective is to offer a satisfactory array of services to consumers. ‘Satisfactory’ of course does not refer solely to product variety. Other dimensions of performance are the overall price level, the price structure and service quality. The third objective is redistribution across consumers and across geographical areas. While the fulfilment of this third objective would ideally not involve intervention at the industry level, various considerations, including prominently informational constraints on income redistribution, have traditionally motivated a manipulation of regulated prices for redistributive purposes. We will come back to universal service obligations in Part two of this report, and now focus on the first two objectives, which appear to be more relevant to a discussion of institutional design. (Note, though, that a move from regulation to antitrust necessarily raises the political transparency of redistribution. The abolishment of regulation leads to the introduction of universal service funds when they did not exist in the regulatory environment. Cross-subsidies among categories of consumers then become apparent to the electorate at large, while in the traditional regulatory environment there are privy knowledge of the industry, its regulator and a few interest groups. We applaud this transparency for various reasons although we acknowledge that it may sometimes encourage a selfish political mobilisation against some desirable redistribution.)

Some readers may find it surprising that competition is not listed as an objective. It is clear, however, that competition is an instrument, not a goal. By encouraging cost effectiveness, innovation, low and properly structured prices, it may contribute to the fulfilment of the objectives listed above.

Industry oversight may strike several rocks. First, the industry overseer, be he a regulator or a competition policy enforcer, typically has imperfect knowledge of the cost and demand structure in the industry. He may therefore find it difficult to limit industry rents while inducing efficiency and innovation, to determine optimal product variety, to put effective but not undue pressure on prices, and to thwart anti-competitive behaviour. Second, the overseer may be captured by (collude with) specific interest groups. Third, the overseer may have limited commitment ability. We will therefore need to assess the possible institutions in view of these impediments to the fulfilment of the goals of government intervention.

## **2. Procedures and control rights**

Antitrust authorities generally assess the lawfulness of conducts. Regulators have more extensive powers and engage in detailed regulation; they set or put constraints on wholesale and retail prices, determine the extent of profit sharing between the firm and its customers (as under cost-of-service regulation or earnings-sharing schemes), oversee investment decisions, and control entry into segments through licensing of new entrants and line-of-business restrictions for incumbents <sup>(1)</sup>. This general picture of a large number of instruments and of potentially high discretionary power held by regulators is of course to be qualified by the many constraints they face in their decision-making: procedural requirements, lack of long-term commitment, safeguards against regulatory takings, constraints on price fixing or cost reimbursement rules (cost-of-service regulation, price caps, etc.), cost based determination of access prices, and so forth.

There is some convergence of regulatory and competition policy procedures. For example in Northern America, regulatory hearings are quasi-judicial processes in which a wide array of intervenors can expose their viewpoints. The enlisting of advocates is prominent in both institutions and is a key factor in the reduction of the informational handicap faced by the industry overseer.

There are however a couple of differences relative to the role of intervenors. In antitrust enforcement, private parties, although they are more constrained in their access to the oversight process, play a bigger role than in a regulatory process. Competition policy officials occasionally conduct independent industry studies, but the vast majority of cases brought to courts are private suits. (The mechanisms of course vary substantially across countries. For example the French Conseil de la Concurrence has a very active early role — possibly as a reaction to a private complaint — in the antitrust process.) Another difference is that interest groups are motivated to intervene in the regulatory process solely by the prospect of modifying policy while they go to court either to modify industry conduct (through a court

<sup>(1)</sup> To be certain, antitrust authorities and courts sometimes exercise regulatory authority by imposing line-of-business restrictions or forcing cost-of-service determination of access prices. A case in point is Judge Greene becoming a regulator of the American telecommunications industry. But the pattern described in the text seems fairly general.

injunction) or to obtain monetary compensation (e.g., treble damages). Yet another difference between the two institutions is that courts have less control over the agenda than regulators. While they can throw a case, they most often examine it and may easily become overloaded. Conversely, they can only take cases which are brought to them.

Another distinction between the two institutions is the separation between investigation and prosecution in antitrust. In contrast, regulators conduct regulatory hearings and adjudicate on their basis. We should however not overemphasise this distinction. First, some competition policy-makers, such as the Conseil de la Concurrence in France, can both investigate and take action against specific behaviours (subject to the possibility of court appeal). Second, regulatory decisions may, too, be appealed in court in the same way a court decision may be overruled by a higher court.

A last point of departure between the two institutions relates to the consistency requirements. Regulators and courts are both required to apply relatively consistent reasoning. But while precedents play some role in the two institutions, regulators are mainly bound to be somewhat consistent with their previous decisions for the industry they oversee. Courts, in contrast, must also refer to decisions of other courts as well as to decisions pertaining to other industries. In particular, the uniformity of interventions across industries imposes substantial constraints on the courts' discretion.

To sum up, and with all the required qualifiers, regulators typically have more instruments and potential discretion than antitrust enforcers. They engage in detailed regulation of prices, investments, rate of return, entry and competition while antitrust enforcers focus on competitive aspects. Regulators also face lighter consistency requirements since they issue industry specific rules.

### 3. Timing of oversight

An important difference between regulation and antitrust is that the former operates mainly *ex ante* and the latter *ex post*. Antitrust authorities assess conduct after the fact while regulators define the rules for price setting, investment and profit sharing *ex ante*. Again, some qualifiers are in order. Merger control by European and American competition policy officials requires notification for large mergers and is a quasi-

regulatory process <sup>(1)</sup>. Conversely, an agency's decision of disallowing *ex post* 'imprudent investments', that is of excluding them from the rate base in a cost-of-service process, is an illustration of *ex post* decision-making in a regulated environment. But the broad picture is that the timing of regulatory decision-making differs from that of antitrust enforcement.

Concomitantly, the regulatory process must be more expedient. The necessity not to halt productive decisions as well as rules constraining the length of investigations often put pressure on regulators (or quasi-regulators such as merger control officers) to converge on rapid decisions. In contrast, the *ex post* nature of antitrust intervention does not call for a similar expediency.

Another implication of the timing of government intervention is that the uncertainty about the overseer's decision-making differs between the two institutions. *Ex ante* intervention removes most of the uncertainty about this intervention (although not necessarily about its consequences). It may thus facilitate financing of new investment by alleviating the lenders' informational handicap with respect to this intervention (to the extent that the lenders may have insufficient expertise in the industry and may thus be concerned about the borrower's superior knowledge about this intervention) and by sharpening the measurement of the borrower's performance (by eliminating extraneous noise not controlled by its management) <sup>(2)</sup>.

*Ex ante* intervention also provides some commitment by the regulator toward the firm. This commitment is desirable whenever the regulator has the opportunity to exploit the firm's demonstrated efficiency or investment by becoming very demanding. We will come back to this expropriation issue later on.

*Ex ante* intervention may be particularly valuable when coordination problems are important, e.g. as for the design of the articulation between urban and intercity

<sup>(1)</sup> See Neven, Nuttall and Seabright (1993) for a very relevant discussion of institutions in the context of merger control. Firms may, but are not required to (except for some licensing agreements), submit vertical agreements for approval to the European Commission.

<sup>(2)</sup> That is, the removal of uncertainty may reduce both adverse selection and moral hazard. Note that the removal of regulatory risk per se does not improve welfare through a reduction in the risk faced by risk averse investors. (To the extent that the regulatory risk in the industry is idiosyncratic it should be diversified away under perfect capital markets.)

transport networks, or between different modes (rail and buses) of urban transport.

Last, *ex ante* intervention may force the firm to disclose information that it would not disclose after the fact. Intuitively, it is less risky for the firm to conceal or manipulate information *ex post* when it knows the state of nature than *ex ante* when it does not; for instance, the firm may know *ex post* that a lie about an *ex ante* information that conditioned some business decision will not be discovered, but it may have no such certainty *ex ante* <sup>(1)</sup>.

A drawback of *ex ante* intervention is that it may foster collusion between the industry and the overseer. The industry knows whom it is facing while it is much more uncertain about whether it will be able to capture the (unknown) overseer in a context in which the oversight takes place *ex post*. This uncertainty about the possibility of capture increases the firm's cost of misbehaving.

A second benefit of *ex post* intervention is of course the opportunity to take advantage of information that accrues 'after the fact'. For example, it may over time become clearer what constitutes acceptable conduct. To be certain, *ex ante* decisions could in principle be flexible enough to allow for *ex post* adjustments that embody the new information; but describing properly *ex ante* the nature of future information that will be brought to bear on the determination of acceptability may prove difficult and not generate much gain relative to a pure *ex post* intervention.

To sum up, regulatory decisions have an *ex ante* flavour while most antitrust decisions are taken *ex post*. *Ex ante* intervention provides several benefits: removal of regulatory uncertainty, higher commitment, and pressure for regulatory expediency. *Ex post* decision-making in contrast is based on information refined by the passage of time. It may also reduce the risk of capture.

#### 4. Information intensiveness and continued relationship

Another useful distinction between antitrust and regulation is that regulatory decisions rely on superior exper-

tise. The regulatory advantage in this respect is three-fold. First, a regulatory agency specialises on a specific industry while antitrust enforcers have a fairly universal mandate. Second, regulators are usually involved in a long-term relationship with regulatees while judges (Judge Greene notwithstanding) are not. Third, regulators usually have larger staffs than judges and monitor the firms' accounts on a continuous basis rather than on an occasional one; they also can insist on specific accounting principles (such as accounting separation) as well as disclosure rules. This information superiority can clearly be more or less important according to the context. It is for instance more likely to be substantial in the case of a single-industry firm regulated by a national agency, as for electricity in the UK or in France, than in the case of a multi-activities firm regulated by local agencies, as for the German *Stadtwerke* or the Italian *aziende*.

Superior expertise is of course a benefit in that it allows better informed decision-making. For example regulators have for a long time used cost-based rules for retail and wholesale prices even though the determination of costs is often a difficult task. Such a task is even harder for a court which has not monitored the firm's accounts in the past and imposed specific accounting and disclosure requirements. It is therefore not surprising that antitrust enforcers are more at ease with cases based on qualitative evidence (price discrimination, price fixing, vertical restraints, etc.) than with those requiring quantitative evidence (predation, tacit collusion, access pricing, etc.).

Superior expertise however may be a handicap when regulators have limited commitment powers. When a regulated firm lowers its marginal cost through efficiency measures or investment, it is tempting for regulators (or politicians) to confiscate the efficiency gains through lower prices. This 'ratchet effect', which is strengthened by the regulator's access to information about the firm's efficiency, is an impediment to efficiency. Similarly, an excessive attention (motivated by superior expertise) may inhibit the firm's initiative. Recent research has shown that an arm's length relationship may entail more commitment power and help provide better incentives <sup>(2)</sup>.

A second drawback of expertise is not due to expertise per se, but rather to the way it is acquired. Part of the

<sup>(1)</sup> Contract theorists would express this in a more technical manner: Incentive constraints *ex ante* are pooled, since they are expressed in expectations. It is therefore easier to elicit information *ex ante* than *ex post*, because there are fewer incentive constraints.

<sup>(2)</sup> See, for example, Crémer (1995), and Aghion and Tirole (1997).



regulatory agencies' expertise stems from the long-term nature of their relationship with the industry. But, as is well-known, long-term relationships are, in any organisation, conducive to collusion. And indeed, regulators have traditionally been more captured by interest groups than judges. This may also be related to the fact that, since regulators have deeper knowledge of a particular industry, a natural career evolution is more likely to involve close links with this industry (that is, the regulators' expertise may well reinforce 'revolving doors' problems). Also, the need for such industry-focused expertise may impose some constraints on the recruitment of regulators.

To sum up, regulators are usually better informed than judges. This has a beneficial impact on government intervention, although it may also jeopardise the firm's incentives to reduce cost. And, expertise, to the extent that it goes together with a long-term relationship with the industry, is conducive to capture by interest groups.

## 5. Independence vis-à-vis the political environment

The final dimension along which we compare regulation and antitrust is their relationship to the political power. Antitrust authorities are traditionally described as being more independent than regulatory agencies. While this view is generally correct, it is important to distinguish among forms of regulation and competition policy: an antitrust authority located within a ministry is more likely to be influenced by politics than an independent regulatory agency.

The Anglo-Saxon institution of regulation by an independent commission seeks to emulate the benefits of an independent judicial system. Independence can be partially obtained by offering long, staggered terms to commissioners, and by limiting the impact of the legislature on the agency's budget and jurisdiction.

The benefits of independence are well-known. First, the politicians' concern about public opinion and their taste for campaign contributions make them prone to give in to interest group pressure. Relatedly, an independent agency may be less sensitive to alternative motivations (such as favouring domestic or public operators), which may reduce regulatory uncertainties and offer a better commitment to fair treatment of all competitors. Examples of such concerns are particularly prominent in

the allocation of airport slots and or rail slots. For instance, French private airlines have repeatedly complained about the allocation of slots, charging the State agency (the Direction Générale de l'Aviation Civile, or DGAC) with favouritism towards Air France and Air Inter. (Euralair, for instance, had at some point authorisation for Toulouse–Orly flights but had no slots allocated to operate such flights.) Independent agencies are less vulnerable to interest groups, although their officers are not immune to the influence of the revolving door and sometimes of bribes; decisions can then be reached more on efficiency grounds and less on the basis of the relative power of pressure groups. This is of course a substantial advantage of independence. Relatedly, independence may strengthen the agency's commitment power by limiting both opportunistic captures of the firms' rents and 'soft-budget constraint' problems. Second, and in the same vein, independent agencies and the judicial system tend to be more professional than executive departments unless strict civil service rules in the latter limit political patronage to a couple of top jobs. Independence also allows for more transparency. In France, for instance, the only European country where the air traffic control is directly managed by the State, through the DGAC, airline companies have argued that the accounting system does not provide a clear enough basis for the fees charged to the companies. Many countries have chosen to give the air traffic control to either an independent agency or a non-profit organisation, and some countries such as the USA are even considering privatising it.

The cost of independence is also well-known. Independent agencies and courts may lack accountability and follow their own agenda instead of the nation's agenda. We should in general take this lack of accountability seriously, but there are reasons to suspect that in the case of the complex oversight of network industries, the public has an especially ill-informed opinion and often no opinion at all. In such circumstances, the public cannot verify whether politicians really defend its interests. (An exception may be the specific issue of the public service. Although the electorate is unlikely to understand the stakes associated with alternative ways of implementing public service objectives, its preferences with respect to broad public service goals should be reflected in public decision-making. On the other hand, universal service obligations may be performed outside the regulatory realm, as will be stressed in Part two, and thus there is no conflict between the establishment of an independent regulatory agency and the fulfilment of politically determined universal service obligations.)

The lack of accountability of independent agencies and courts have always been a motivation for the introduction of constraints on their discretion: procedural requirements, limited discretion and commitment power, and possibility of appeal. This observation however does not negate the point that their decisions are less influenced by interest groups than those taken by executive departments.

There has been remarkably little work in economics on the costs and benefits of independence. Lenoir's (1991) model exhibits a trade-off involved in making an agency independent from its political principal. Her stylised model depicts a three-party hierarchy: political principal (Congress, or more realistically Congressional committees and subcommittees in charge of overseeing the industry), regulatory agency and the industry (e.g., a monopolist). The political principal can adjust the appropriations received by a dependent agency to the latter's real needs, while an independent agency's budget is protected from political intervention. That is, the benefit from having a dependent agency is the possibility of tailoring the agency's resources to the specific circumstances. Thus, Lenoir's model focuses on a particular version of accountability — that the agency not waste resources — but various other versions of the concept of accountability, which more generally stands for the alignment of the agency's behaviour with the public interest, would be consistent with the overall argument.

The cost of dependency in Lenoir's model is the influence of politics on regulatory decisions. The influence of the interest group (the industry) on the regulatory agency flows through the political principal. Namely, the industry can offer campaign contributions to members of Congressional committees and subcommittees in charge of overseeing the agency and thus the industry. The political principal has no control over an independent agency and thus the industry cannot influence regulation through the political system. In contrast, the political principal can threaten to reduce a dependent agency's budget and thus to affect its rent; it can then offer not to ratchet down its budget to the efficient level in exchange of the agency's lenient treatment of the industry. Thus, a dependency relationship creates a *quid pro quo* and allows the industry to impact indirectly on regulatory decisions <sup>(1)</sup>.

<sup>(1)</sup> This is an illustration of the more general point that collusion is enhanced by a mutual power relationship: see Laffont and Meleu (1996).

We are not aware of much research on agency independence in political science either. Fiorina (1985) focuses on other determinants of independence than the efficiency ones reviewed here. In his view the independence of an agency is not necessarily determined by whether an independent agency produces a more efficient outcome than a dependent one; rather it relies on the legislators' desire to shift or take responsibility. Delegation enables politicians to shift the blame but prevents them from claiming credit for policies. There is of course much work in political science on the working of government agencies and ministries in general (Bernstein 1955, Yates 1982, Wilson 1989). Of particular interest for this study is Bernstein's theory of the life cycle of regulatory agencies. Bernstein contends that the life cycle of an independent agency can be divided into four periods: gestation (production of a regulatory statute), youth (lack of experience, unclear mandate, creative, aggressive and crusading spirit), maturity (focus on the protection of the agency's existence and power, switch from policing to managing the industry, higher concern with the health of the industry, loss of vitality, desire to avoid conflicts), and old age (extreme conservatism, loss of creativity, growing backlogs and apathetic approach to the public interest in regulation). While Bernstein's theory necessarily vastly oversimplifies reality, it relates to some of the themes we have described earlier, such as the impact of continued relationships.

## 6. Toward the demise of regulation?

Prior to the liberalisation of network industries, antitrust could not have contributed much to the control of vertically integrated monopolies. The most obvious handicap of antitrust policy is that it is mostly concerned with the fairness of competition (besides consumer protection). But competition was rarely to be seen. Regulators were often averse to it and did their best to preserve the status quo. And even in the absence of regulatory restrictions on entry, economic obstacles stood in front of prospective entrants. (Entrants are handicapped by the early-mover advantages of incumbents: 'Stackelberg' capital accumulation in industries with large investments, on the one hand, and, in the presence of network externalities, existence of an installed base of customers that either needs to be eroded in order for the entrant to offer a decent network size or requires the writing of an ironclad interconnection agreement.)

The view that regulation was the proper institution in the pre-liberalisation environment is further strengthened by our comparative discussion of the two institutions. The control of monopoly prices and profits requires high intensity monitoring, control rights to regulate prices and returns, and a combination of *ex ante* oversight and guarantees against regulatory takings that provide some assurance to the firm that its efficiency gains and investments will not be expropriated. The antitrust institution does not exhibit any of these features.

Two main factors contributed to the recent liberalisation reforms worldwide. In some network industries, technological change has reduced the importance of returns to scale and enhanced the prospects for competition. And mainly the dissatisfaction with regulatory performance led many decision-makers and economists to the view that competition would exercise a better control over incumbents than poorly informed or captured regulators.

The advent of competition was bound to call the existence of regulation into question. For true believers in market theory, competition will force prices down, quality up and profits to zero. The very objectives of regulation listed above will therefore be fulfilled; regulation will be at best superfluous and perhaps even nefarious to the extent that it attempts to interfere with the market mechanism. Antitrust enforcers need only check that this mechanism is not derailed by unlawful conducts.

While anyone versed in antitrust theory or practice would question the relevance of such bold statements to most industries, there are specific reasons to believe that traditional, stand-alone antitrust enforcement may in most network industries not provide the smooth competitive environment it is supposed to create. A key feature of a number of network industries is the need for mutual interconnection among competitors. While it is straightforward to mandate open access on paper, it is much harder to confront the subtle issues involved in developing compatibility and interconnection: technological requirements, level of unbundling, quality and timing of interconnection and level of interconnection charges. Each of these dimensions can give rise to anti-competitive foreclosure behaviours by incumbents on the one hand, and to excessively costly demands by entrants on the other hand. Designing a proper interconnection policy requires not only (i) a sophisticated

understanding of economic incentives and effects, but also, and to a varying degree <sup>(1)</sup> depending on the regulatory mode, (ii) substantial technological expertise, and (iii) considerable cost and demand information. Although courts occasionally investigate cases of alleged foreclosure by an essential facility owner in unregulated environments, they are unlikely to engage in detailed oversight of interconnection policies. And indeed, as illustrated by the Clear-Telecom dispute in New Zealand, courts have been somewhat reluctant to become involved in such policies.

The *ex post* nature of competition policy also creates considerable uncertainty for entrants who have to sink substantial investments without knowing an important determinant (the interconnection charge) of their profitability. This further handicap of competition policy in the matter of network industry oversight is even heavier when the entrants need to leverage themselves up in order to finance the said investments, since the capital market may be concerned by the enhanced riskiness of lending.

The *ex ante* determination of the conditions of interconnection can occur in one of three settings. Under *laissez-faire*, firms arrive voluntarily at a mutually beneficial agreement. The *laissez-faire* policy raises two concerns that will later be elaborated upon. First, in the liberalisation phase of the industry, incumbents may deny proper access to entrants (foreclosure concern). Second, in the more mature phase, firms may agree on high access charges in order to raise final prices (collusion concern). Under compulsory arbitration, firms in case of a dispute must accept the decision of an arbitrator whom they or the government have previously selected. Last, regulation can set the rules for the determination of interconnection charges, if not these charges themselves.

It is at this stage worth discussing the possibility of using arbitration in interconnection disputes. While arbitration is routinely used to settle disputes in a number of areas (labour contracts, long-term procurement contracts, or international transactions), this interesting alternative has not been thoroughly studied in the con-

<sup>(1)</sup> In the context of one-way access, we favour a light-handed regulation in order to avoid a situation in which regulators attempt to manipulate access prices and thereby create an antagonism between the interconnecting parties. See in particular Laffont and Tirole (1996a, b).



text of network industries (except in Australia and New Zealand). Arbitration raises the prospect of an expedient, politics-free determination of interconnection charges. Its application must however be preceded by an analysis of several questions. First, while it alleviates the first concern listed above (foreclosure concern), it does little to tackle the collusion concern since firms can opt out of arbitration and agree on an high mutual interconnection charge. Second, the agreements reached between an incumbent and new entrants are conditioned by what they believe will occur if they go to arbitration, and therefore by the instructions received by the arbitrator on how to resolve the dispute. That is, the arbitrator's 'regulatory rule' for determining access prices (or choosing among the parties' final offers under final offer arbitration) is key to the outcome of interconnection negotiations. Third, there is some question as to who should play the role of the arbitrator. Ideally, the arbitrator should be an independent and benevolent expert. Our discussion however points at a trade-off between expertise, which is often associated with a long-run relationship with the industry, and independence, which is jeopardised by frequent interaction with the industry's interest groups. The discussion also raises the question of the overseer's timely access to detailed information (<sup>1</sup>).

While we need to think about new and innovative modes of access price determination, regulation is likely to remain one of the institutions of industry oversight. This does not mean that regulation will not exhibit the drawbacks it has been derided for. In particular, the pressure of interest groups will not subside with the advent of competition. While the traditional slant of regulation for decades has been an overprotection of incumbents, one can also imagine that the

widespread and historically founded suspicion of regulatory capture by incumbents as well as personal ambition (of attaching one's name to the liberalisation of an industry) will in some instances lead regulators to 'overshoot' in their encouragement of entry and to manage entry so as maximise the entrants' market share to the detriment of efficiency. Partisans of regulation assert that regulators will quickly strike the right balance and focus on efficiency rather than on private interests while sceptics argue that they will always be prone to favour specific interest groups on the basis of their relative power and of the political agenda. We are sympathetic to the views of the sceptics, but we of course plead for the use of sound economic reasoning as a partial safeguard against politics; and we hope that future research will develop better conceptual frameworks for antitrust in the context of network industries.

## **7. Are antitrust and regulation substitutes or complements?**

Are we bound to choose between antitrust and regulation or should we opt for a dose of each?

The case for substitutability is based on the theory of multiprincipals. There is a large literature in economics concerned with the inefficiencies associated with split oversight of agents. These inefficiencies are not a pure construct of the mind. Practitioners of federal decision-making know that the coordination of policies (for example, regional, national and supranational) is by no means an easy exercise. Examples also abound in the regulatory sphere. The lack of coordination between the Environmental Protection Agency and the State-level Public Utility Commissions in the USA has created difficulties for the participation of power companies to newly created markets for SO<sub>2</sub> pollution permits. Likewise, the separation between the prerogatives of the Federal Communications Commission and the Public Utility Commissions creates problems of consistency and incentives in the regulation of the regional telecommunications companies with respect to the determination of retail and wholesale prices and of profit sharing. Similarly, future theoretical research and experimentation should shed some light on whether the coexistence of regulators and courts cannot create problems of coherency of the oversight process.

<sup>(1)</sup> One of the advocated merits of arbitration is to limit direct intervention to the (supposedly exceptional) cases that cannot be resolved directly by the parties, while at the same time providing strong incentives for the parties to reach agreement through commercial negotiation rather than to resort to final arbitration. However, some commentators have questioned this approach. For instance, small access seekers may be unwilling or unable to take a dispute to arbitration (e.g. because of informational disadvantages, or of concerns about retaliation by the facility operator) and have no choice but accepting a monopoly access price. Also, the infrastructure owner has an incentive to bribe downstream users of the facility to accept the monopoly price, in return for a share of the resulting monopoly rent (that is, arbitration may lead more to a sharing of monopoly rents than to their dissipation). In both cases, those commentators observe, access disputes would be rare but that would not necessarily be an indication that the regime had successfully eliminated monopoly pricing.

The case for complementarity is threefold <sup>(1)</sup>. First, having multiple overseers may reduce the scope for capture <sup>(2)</sup>. Concretely, the antitrust process may provide a forum for entrants when the regulator is excessively lenient with the incumbent. Second, the separa-

tion of tasks may create a more focused mission to each overseer. While the desirability of creating advocates for various causes is a general principle of government design <sup>(3)</sup>, more thought has yet to be given to the application of this general principle to the interaction between antitrust and regulation. While it is natural to think of antitrust authorities as competition advocates, their exact role in the determination of interconnection prices is still to be determined. Third, the *ex ante* nature of regulation and the *ex post* nature of antitrust creates a nice division of labour between decision-maker and assessor. As is well-known in some institutions (e.g., banking regulation), the confounding of these roles can easily create incentives for cover-ups.

Overall, we suspect that regulation and antitrust have a rather complementary role, but we also acknowledge that the state of economic thinking in the matter is too rough to bring a definitive answer.

---

<sup>(1)</sup> We focus here on the complementarity of institutions. Another notion of complementarity is the complementarity between *ex ante* and *ex post* oversight. The recent Australian regulatory reform in the telecommunications industry may illustrate the distinction between these two notions of complementarity. Before the reform, the Australian Telecommunications Authority (known as Austel) was under the Telecommunications Act 1991 the specific regulator in charge of the industry (together with the Spectrum Management Agency). The Telecommunications Bill of 1996 and the Trade Practices Amendment Bill of 1996 merges those two agencies to create the Australian Telecommunications and Spectrum Management Authority (to be known as Austel again), but restricts its scope of intervention to technical issues (such as setting codes of practices) and transfers the 'competition policy regulation' (including price supervision for monopolistic segments) to the competition authority (the Australian Competition and Consumer Commission). So, the antitrust authority exercises *ex ante* control (on monopoly segments) and *ex post* control (on competitive segments). In some other Anglo-Saxon countries (US, UK), regulators have kept more prerogatives relative to *ex ante* oversight. Moreover, in the UK the regulator (OFTEL) plays a substantial role in the *ex post* oversight of the telecom industry.

<sup>(2)</sup> See Laffont and Martimort (1995).

---

<sup>(3)</sup> See Dewatripont and Tirole (1995) for the underlying theory.

## IV. Pricing interconnection

### Introduction

The analysis of access (interconnection) is largely based on the ongoing debate in the telecommunications industry. The conceptual analysis of access in other industries (electricity, railroads, etc.) is less developed. Although it shares a number of features with that of telecommunications, the reader should be aware that other industries have specificities that often have not yet been accounted for in the literature.

Historically, infrastructure monopoly owners also have had a monopoly position on the complementary segment (services). Interconnection between services and the network has been an internal technical matter for the monopoly, with the regulator regulating only the final products <sup>(1)</sup>. The advent of deregulation has raised a number of complex issues concerning the access of competitors on the complementary segment to the infrastructure.

We consider two paradigms. In the first (case 1), the infrastructure (or network) is a monopolised essential facility. The focus is on ‘one-way interconnection’, that is on the interconnection of competitive service providers to the essential facility. In the second paradigm (case 2) the infrastructure has been duplicated, and network externalities are reaped through ‘two-way interconnection’.

In case 1a, the infrastructure operator is still vertically integrated but competes with other service providers. These competitors on the complementary segment must be connected to the network. This is the most prominent

case today as competition is introduced in formerly State owned monopolies industries. Section 1 is devoted to this set up. In case 1b, the infrastructure operator is prevented by line-of business restrictions from entering the complementary segment. Such structural separation for example prevails in the (pre-1996) USA telecommunications industry and in the UK power industry. The pricing of interconnection is simpler since the network operator cannot attempt to favour its own services. The theory of interconnection is then a degenerate case of the theory developed in Section 1.

The choice between 1a and 1b in a regulatory environment corresponds to a trade-off between economies of scope (that are often difficult to measure), and the probability that the regulator cannot prevent the network operator from favouring its own services. Since perfect regulation can always mimic structural separation, a non-trivial analysis must introduce regulatory imperfections in order for there being a trade off in the choice between vertical integration and vertical separation. Section 2 analyses the incentives for foreclosure by the network operator in the extreme case of a deregulated environment.

Section 3 illustrates interconnection issues under vertical integration and structural separation through recent reforms in the power and gas industries in the UK and the USA.

In case 2, the case of ‘two-way access’, the networks have been duplicated but are interconnected and the consumers connect to one of them. This is already the case for telecommunications in some countries like the UK, Sweden and soon all over Europe. Section 4 is devoted to this important case for which no economic analysis was available so far.

<sup>(1)</sup> See Laffont and Tirole (1993) for a theoretical treatment of the regulation of natural monopolies. Our treatment in Section 1 borrows from Laffont and Tirole (1996b).

## 1. Pricing interconnection of competitive services to a monopoly infrastructure network

### Introduction

#### Background

A major policy issue in network industries is the liberalisation of potentially competitive segments which need the network as an essential input, which is thus an ‘essential facility’ or ‘bottleneck’. The essential facility is often monopolised because of large economies of scale, of first-mover advantages or of technological superiority. The policy-maker must induce an efficient allocation of resources. This involves, among other things, creating proper conditions for entry into the competitive segment while not inducing excessive entry, not expropriating previous investments or discouraging future ones in the monopolised segment, or not generating inefficient bypass.

This question is a classic one in antitrust economics, as courts have been asked to investigate the existence of foreclosure and the design of relief policies in industries as diverse as the cement, railroad and computer reservation systems industries <sup>(1)</sup>. Also, and it is the subject of this section, it has received substantial attention lately in the regulatory context of the creation of competition in various segments of the activity of a dominant natural monopoly. The impetus for the development of interconnection policies in the telecommunications industry was the opening of competition in the long distance markets. New entrants (such as Mercury in the UK and Sprint and MCI in the USA) needed access to the dominant operators’ (British Telecom, AT&T) local networks to reach the customers. It then became clear that interconnection charges would play a much broader role as competition started developing in the local segment from cable companies and mobile operators who needed access to long distance services and in value-added services, and as networks began to proliferate <sup>(2)</sup>. Neither is the issue specific to telecommunications. It has been argued for instance that transmission for electricity, pipelines for gas, tracks, rails and stations for railroads, and mail collection and distribution for post

offices are natural monopoly segments to which rivals in complementary segments (generation of electricity or production of gas, freight or passenger train operation, mail transportation) must have access in order to compete.

Policy-makers have taken one of two approaches <sup>(3)</sup>. The radical approach has been to break up the vertically integrated dominant firm and to prohibit the essential facility spin-off from re-entering the competitive segments. In 1984, in what is perhaps the most famous recent divestiture (resulting from the 1982 Modification of Final Judgment), the American local exchange telephone companies were split from the long-distance operator AT&T, and forbidden to enter the long distance and value-added services markets <sup>(4)</sup>. The American Department of Justice and the Court felt that it was too difficult for a regulator to create a level playing field on the competitive segments. Similar divestitures have also been undertaken in the British electricity system and for the British and Swedish rail systems. In air transportation, the two most important bottlenecks are the airports and the air traffic control, which are both typically separated from airline companies.

The divestiture approach is often criticised for ignoring economies of scope or for being cumbersome, especially in environments where rapid technological changes affect the location of bottlenecks. Neither does it fully solve access allocation problems, as testified by the never-ending battle over airport slots, nor the determination of access charges, although the symmetry among competitors certainly facilitates it <sup>(5)</sup>. A more common policy consists in preserving the dominant vertically integrated firm while regulating access prices so as to create a level-playing field and promote competition. The Anglo-Saxon countries (USA, Canada, UK, Australia, New Zealand) have intensely debated the merits of various interconnection rules, and the European Commission has issued a Green Paper on the

<sup>(1)</sup> See Section 2 for an overview of the theory and applications of the foreclosure doctrine.

<sup>(2)</sup> See Hausman (1994a) for an overview of network proliferation in telecommunications.

<sup>(3)</sup> The regulators’ dual approach in network industries only echoes decisions in industries subject only to competition policy. There have been break-ups, as well as interconnection policies as in the well-known American 1984 CAB decision on computer reservation systems.

<sup>(4)</sup> The separation is in fact not complete. For example, local exchange carriers supply access to the local loop to inter exchange carriers with which they compete in intra LATA toll services. Besides, RBOCs will soon be allowed to enter the inter LATA and manufacturing markets, while AT&T will be allowed to offer local services.

<sup>(5)</sup> Furthermore, competitors on a given segment (say, a local loop) need mutual access in the presence of network externalities (unless all consumers subscribe to all networks). This issue is dealt with in Section 4.

liberalisation of telecommunications infrastructure and cable television networks.

### *Interconnection policies*

The recent years have witnessed a broad array of recommendations and practices. Entrants typically argue for a cost-based access charge such as long-run incremental cost of access. Although this approach was adopted in Australia for telecommunications, it is widely accepted that marginal cost pricing of access prevents the dominant telephone operator from efficiently recovering the fixed costs of the network (and possibly the deficit stemming from the universal service constraints.) Many regulators and economists have suggested long-run incremental cost plus a mark-up as a workable alternative to allow recovery of the 'access deficit'. What is put into the mark-up is the object of intense debate. Should it be an accounting allocation of the access deficit as under fully distributed costs? If so should it take the form of a uniform mark-up on all services or else of a mark-up proportional to long-run incremental cost (the 'Allais rule'), as was suggested in the influential WIK-EAC report written for the European Commission <sup>(1)</sup>? Or should the mark-up be related to the use that is made of the access by the telephone operator's competitors and therefore depend also on demand considerations?

Proponents of purely cost-based and usage-based mark-ups have been arguing for years, with the usage-based approach recently scoring points after a slow start. Two 'potentially <sup>(2)</sup> usage-based' rules have been adopted. In the UK, the Office of Telecommunications (OFTEL) has implemented a practical rule that links access charges to the loss in British Telecom's profit on the segment <sup>(3)</sup>. Even more recently (October 1994), the New Zealand Court of Appeal, following the lead of the California Public Utility Commission, explicitly endorsed the 'efficient component pricing rule'

(ECPR), also called the 'Baumol-Willig rule' <sup>(4)</sup>, as being the most efficient approach allowing Clear Communications Limited to compete with the incumbent Telecom Corporation of New Zealand <sup>(5)</sup>. ECPR picks an access price equal to the difference between the network operator's price and marginal cost on the competitive segment. That is, it charges a competitor an access price equal to the margin lost by the network operator on the competitive segment.

There are really two main questions. What is the theoretically correct benchmark? And how can one translate theoretical precepts into workable recommendations? While cost-based rules are advocated on practicability grounds, usage-based arguments derive their legitimacy from a theoretical approach. For example, both Baumol and Willig have taken the view that ECPR is the logical implication of the theory of contestable markets. We will strongly concur with the idea that usage-based rules are the proper theoretical benchmark (and will of course discuss the practical difficulties in implementing such rules). Section 1.2 develops a theoretical framework in which to assess cost and usage-based rules.

On the theoretical front, ECPR has some appeal as we will see. It is however no panacea, for three reasons. First, the contestable markets paradigm as such seems of limited value for studying access pricing as it predicts that the competitive segment will be monopolised either by the network operator or by its competitors <sup>(6)</sup>. Second, it is only a partial rule as it does not specify how to determine the network operator's prices on the competitive segments (which form the basis for the computation of the access prices) <sup>(7)</sup>. This is no small issue as we will see. Our third quibble with ECPR

<sup>(1)</sup> The WIK-EAC report calls for some downward flexibility of interconnect prices from the level given by long-run average-incremental cost plus equal mark-up. The motivation for downward flexibility is to enable the operator to compete with bypass.

<sup>(2)</sup> 'Potentially' refers to the fact that these rules only link access prices and final prices and that, therefore, their content depends on what policies are adopted for final prices. As we will see, the rules are actually not usage-based under certain regulatory modes.

<sup>(3)</sup> It should be noted, though, that the December 1994 and December 1995 Consultative Documents issued by the UK Director-General of Telecommunications ponder a move towards (an approximation of) incremental cost pricing of access (with small, equal mark-ups).

<sup>(4)</sup> This rule was first proposed by Robert Willig (1979, pp.140 and 149), popularised by William Baumol in numerous regulatory proceedings and in writings (Baumol (1993), Baumol and Sidak (1994b)), and endorsed by several other prominent US economists (e.g. Alfred Kahn and William Taylor (1994)). It is also called by various authors the 'principle of competitive parity', the 'imputation rule' or the 'absence of vertical price squeeze'.

<sup>(5)</sup> The judgment of the Lords of the Judicial Committee of the Privy Council, delivered the 19 October 1994, provides a detailed description of the arguments and concludes that 'in the end, it is (the High Court's) judgment that implementation of the (Baumol-Willig) Rule is more likely than the alternatives to improve efficient competition in New Zealand telecommunications'.

<sup>(6)</sup> As Ergas and Ralph (1994) point out, the contestable market paradigm does not really explain why regulators encourage competition in the first place. That is, the paradigm abstracts from the very elements (differentiation, entrants' superior technology, yardstick competition, or regulatory capture) that make competition desirable.

<sup>(7)</sup> See, for example Baumol and Sidak (1994b) and Baumol and Willig's joint brief in the recent New Zealand judgment for clear discussions of this.



applies to alternative policy recommendations as well: in our view, it makes limited sense to propose a general access pricing rule without consideration of the environment in which access is provided.

### *Conceptual framework*

We will use a straightforward framework in which competition is motivated by product differentiation, and/or cost differences, and/or benchmarking, and in which the network operator is optimally regulated and its competitors face no entry cost, have no market power and cannot bypass the essential facility. This will yield a first assessment of prominent interconnection rules.

Reality however rarely fits this idealised framework. Once bypass, barriers to entry, entrants' market power, and universal service requirements for the network operator are brought into the picture, it becomes very important to consider the panoply of instruments available to the regulator: are the network operator's final prices regulated using cost of service regulation, price caps, market share objectives, or a simple antitrust approach (New Zealand Telecom is only subject to the general prohibition on the improper use of a dominant position)? Is the telephone operator allowed to perform second- and third-degree price discrimination? Can the regulator prevail herself/himself of various 'fiscal instruments' to control the competitors' final prices, such as a tax on their outputs, of taxes on bypass, of proceeds from a universal service fund, or of entry subsidies? If so can the regulator allocate these taxes and subsidies to the network operator or to the competitors? These questions derive their importance from the fact that, as the number of market imperfections grows, the access pricing rule is bound to respond to an increasing number of concerns, *ceteris paribus*, and that more instruments are needed if access prices are not to arbitrate inefficiently among conflicting goals.

### *Workability of the optimal access pricing rule*

The policy debate appropriately devotes much attention to the practicability of existing proposals. For example, applied literally to the telecom industry, ECPR requires measuring a telephone operator's cost savings on each of its thousands of competitive segments (defined by routes and services) <sup>(1)</sup>. On the other hand, proponents

of long-run incremental costs with or without proportional mark-ups, who often deride ECPR on practicability grounds, must confront the fact that their proposal requires measuring a huge number of access costs. Accordingly, we will discuss informational requirements in detail.

We offer for consideration an alternative simple policy for interconnection. This policy's main attraction is that it follows the theoretical precepts. And, while it imposes some informational demands, it requires no more information than existing schemes. The policy consists in regulating the owner of the essential facility according to a global price cap. A global price cap includes both access charges and final goods prices. By decentralising price decisions, including those relative to access, a global price cap implements the optimal Ramsey price structure (that is, the prices that minimise the social cost of charging prices differing from the services marginal costs in order to cover the firm's fixed costs) conditional on the firm's knowledge about its demand and cost structures and does not require the regulator to measure marginal costs or estimate demand elasticities.

The key insight is that the inclusion of access prices in the price cap re-establishes the symmetry between access goods and final goods, and partly reconciles the firm with the existence of competition. The firm is led to view its competitors' output as an output of its own, that it partly produces (in the bottleneck segment) and partly outsources (in the competitive segment) if it is efficient to do so. We put forth two versions of our global price cap (others should be considered as well): the plain version, and an enriched version in which ECPR defines a ceiling on access prices. We compare the properties of these two versions.

### **1.1. A theoretical framework**

For illustrative purposes we use the following paradigm: a network operator or monopoly (M) fully controls a network and faces competition (C) by one or several competitors in the provision of services using the network as input.

For expositional simplicity, all activities exhibit constant returns to scale, except for the existence of a fixed cost in the network. This fixed cost which will represent the access deficit broadly defined in this framework, can be thought of as having two components (we will later argue that other components may be added to this

<sup>(1)</sup> See, for example, Ralph (1994).

access deficit). The first is the set up cost of the network. The second is the difference between a universal service deficit (linked with the constraints of creating easy access to phones and of national uniform pricing) and the fixed subscription charges paid by the consumers, unless a universal service fund explicitly covers this difference.

An access charge is paid by the competitors to the monopolist. We will assume in a first step that competitors have no individual market power. We will later explain how the access rule must be modified in the presence of market power.

*Remark on the nature of the access deficit.*

A central issue in interconnection policy is that the covering of the access deficit (broadly construed) creates substantial misallocation of resources. It is therefore worth trying to reduce this access deficit. For example, in the telecom industry in Europe, one could eliminate the general cross subsidy of local service to all households and replace it by targeted subsidies (such as 'life-line services') to low-income customers and for high-cost areas <sup>(1)</sup>. One could also raise the subscriber line charge whenever this does not induce customers to disconnect <sup>(2)</sup>. While this section takes the access deficit as given and analyses the optimal way to cover it, it is clear that attention must also be paid to the size of the access deficit.

<sup>(1)</sup> See, for example, Hausman (1994a,b), who suggests that all providers of telecommunications services contribute to a 'Universal Service Fund' to provide the targeted subsidies.

<sup>(2)</sup> Whether one can indeed eradicate the access deficit by using high, distortion-free subscriber line charges is a matter of intense debate, and, in our view, is an empirical question whose answer is likely to be time- and industry-dependent. Three factors may make access deficits hard to eradicate. First, the elasticity of demand for subscription may not be negligible. While few customers forego telephone service altogether for small increases at the current subscriber line charges, they may forego the use of multiple lines at home or in secondary dwellings. And, more and more, they may start bypassing the network (through mobile phones). A proper computation of the elasticity of demand must also take into account the loss in revenue from phone users who would have called the disconnected customer. Second, and relatedly, there may be constraints in the pricing of the connection charges. For instance, bringing optic fibre or an electricity cable into a remote village may have some features of a public good, so that efficient pricing is unlikely to cover the corresponding fixed cost. Third, there may exist political constraints. As an illustration, consider the argument that some of the customers who would disconnect for a high subscriber line charge are people who badly need access to a phone (to be able to call a doctor or relatives); the proper reaction to this argument is that one can offer 'menus' of telephone contracts, including some with very low subscriber line charges and very high marginal prices for calls beyond some low volume (as is sometimes done). Politicians may however be reluctant to adopt such policies.

Let us review three rules and introduce a new one. We assume that the regulator possesses the relevant information to implement them. This obviously may require careful audits <sup>(3)</sup>.

## **1.2. Fully-distributed costs**

The most commonly used approach consists of allocating joint costs according to some mechanical accounting rule. There are of course an infinity of potential such rules <sup>(4)</sup>.

For example, one can allocate fixed costs proportionally to the consumption that is made of the local network. That is, a uniform mark-up is added to the marginal costs of the two final goods and of the access good produced by the monopolist. In the case of output-proportional mark-ups, the total benefit covers the fixed cost, so that the monopolist's budget is balanced. Alternatively, one could impose price- (or marginal-cost-) proportional mark-ups where the coefficient of proportionality is chosen so as to satisfy the budget constraint. This rule is also known as the 'Allais rule'.

Fully-distributed-cost pricing has been as frequently decried by economists as it has been used in practice. Its flaws are well-known. First, fully-distributed-cost pricing is cost based and therefore does not encourage cost minimisation. Second, it yields an improper price structure, and is far from being an optimal way of financing the access deficit, even if one abstracts from incentive considerations. Because it is cost based, it 'subsidises' inelastic demand segments to the detriment of elastic demand ones. Similarly, fully-distributed-cost pricing lacks the flexibility needed to deal properly with large customers through non-linear tariffs; this lack of flexibility yields an inefficient allocation of resources at the access and final goods levels in the absence of competition, and creates inefficient bypass in its presence. Relatedly, some forms of fully-distributed-cost pricing create inefficient entry in the competitive segment.

The next two rules define ways to force entrants to internalise the network operator's opportunity cost,

<sup>(3)</sup> See, for example, California's Public Utility Commission's decision (1994, pp. 222–225) for a discussion of the difficulty of computing price floors on the competitive segments when costs on various segments are bundled.

<sup>(4)</sup> See Mitchell-Vogelsang (1991, pp. 137–140) for a review of fully distributed-cost pricing and of its drawbacks.



namely the profit foregone on a competitive segment by the network operator when entrants gain market share on that segment. Both are usage-based in that they lead to access prices that are contingent on the particular use of access that is made by competitors.

### 1.3. The OFTEL rule

The British Office of Telecommunications (OFTEL) has designed an original interconnection policy for British Telecom (BT) <sup>(1)</sup>. Start from the access deficit (AD) to be covered by mark-ups. This access deficit in our context corresponds to the fixed cost of the local monopoly.

The OFTEL rule is a usage-based rule. The competitors pay a 'tax', or 'mark-up', or 'access deficit contribution' (ADC) on a call proportional to the profitability of that call for British Telecom on each line of product.

- All benefits depend on the access price. The access price is then the outcome of a fixed point process. In practice, the access deficit contribution can only be based on historical data or on forecasts. (This remark also applies to the other methods discussed in this part, whether cost based or usage based.)
- In the case of multiple competitive segments this approach yields differentiated access prices, unlike fully distributed costs, which define an access deficit contribution that depends only on the use of the network by the competitors and not on the nature of the products supplied by them. Furthermore, the products are regulated according to a price cap <sup>(2)</sup>. The basket of goods subject to this price cap includes the final goods, but not the access good, whose price is determined separately by the above formulas. We will call such a price cap a 'partial price cap'.
- The OFTEL rule takes a very simple form when the network operator's budget is balanced. Under budget balance, access prices are exactly equal to the

operator's 'opportunity costs'. In other words, the network operator's external price for access is equal to a notional internal transfer price computed by assuming that its competitive divisions charge their marginal cost (including the internal transfer price on the competitive markets). Under budget balance, the OFTEL rule boils down to the efficient component pricing rule.

The December 1995 OFTEL consultative document envisions a more hybrid reform for the UK. Interconnection charges are split into two categories. First, the regulation of terminating access would occur at a very desegregated level. It would be based on marginal costs and would therefore not reflect demand considerations. Second, BT would face a price cap on the basket of all interconnection services (origination and termination), thus giving the operator more freedom to affect the structure of relative access charges at the originating end. The rationale for distinguishing origination and termination is that origination is more easily bypassed and therefore more competitive. The reform in the UK also distinguishes two categories of retail services: those deemed competitive which, like in the USA, would be left unregulated, and those for which competition is still emerging, which would be subject to a separate price cap.

### 1.4. The efficient component pricing rule

As we have seen, the efficient component pricing rule equates the access price to the difference between the network operator's price and marginal cost on the competitive segment, or equivalently to the sum of the net benefit of providing the competitive good and the marginal cost of giving access.

This rule has been generalised to allow for differences in interconnection costs <sup>(3)</sup>. Baumol, Willig, and others have repeatedly stressed that the proper yardstick for defining access prices is that of the perfectly contestable market and that, when supplying an intermediate good to another firm, a supplier must be permitted to price the intermediate good at a level sufficient to compensate it for the sacrifice in profit due to the supply to the other firm. The expected social benefits of charging the opportunity cost for interconnection are

<sup>(1)</sup> The following schematises the policy in order to better highlight its main features. See Cave (1993) or the Director-General's statement on 'Policy on Separation and Interconnection' (1992) for more details.

<sup>(2)</sup> Subject to some further constraints (e.g. the RPI+2 constraint on rentals that slows down rate rebalancing.).

<sup>(3)</sup> Hausman (1994a), California PUC (1994, pp. 213–214).

(see, for example, Baumol and Sidak (1994)). (a) The rule sends the right signal to entrants. Potential entrants can enter profitably if and only if they are more efficient than the network operator. (b) Entry is neutral regarding operating profit for the incumbent operator. This has several consequences. Entry on the competitive segment does not interfere with the cross-subsidisation of the bottleneck segment by the competitive segment; this is a powerful political argument as politicians are very eager to maintain these subsidies. Revenue neutrality also reduces the network operator's incentive to destroy the level playing field by skimping on the quality of access. These two properties have deservedly attracted attention, and certainly have much contributed to the popularity of ECPR.

The conceptual framework advanced to defend ECPR — contestable market theory — makes strong assumptions and yet does not provide completely convincing foundations for the necessity of the ECP rule. Standard perfectly contestable market theory (Baumol, Panzar and Willig (1982)) assumes that all firms (incumbents and entrants) face identical cost functions and the same demand function in the competitive segment. Because it also abstracts from incentive issues, it provides no reason for having more than a single firm in the competitive segment in the absence of diminishing returns in that segment. So in the standard framework of cost symmetry, there is no clear motivation for entry. Proponents of ECPR therefore must have in mind an extension of perfectly contestable market theory in which entrants are (at least sometimes) more efficient than the incumbent operator in the competitive segment. On the other hand, ECPR implies that the network operator in equilibrium supplies only access and exits the competitive segment. This does not mean that the access price is completely irrelevant, because it defines who (the network operator or competitors) will corner the potentially competitive segment. Rather, the prediction of the rule is too stark.

We do not view this argument as building the most important case against ECPR. The unpalatable property just described can be eliminated as shown below. Rather we build on their useful insights to develop a reference model delivering actual competition on the competitive segment and allowing a first assessment of ECPR, before embarking the arduous task of adding various real world distortions. To this purpose, we maintain as much of the spirit of the ECPR argument as possible. We introduce product differentiation as a

motivation for competition <sup>(1)</sup> and allow cost and demand asymmetries between the telephone operator and the competitors. Our normative framework allows us to answer two key questions linked with ECPR:

- (1) What is the proper price reference for the computation of the access price?
- (2) Is ECPR (or its extension to allow for differences in interconnection costs) optimal?

*Baumol and Sidak's precept for the network operator's final price.*

As is well known, ECPR is only a partial rule. It is consistent with many pricing models, including cost based models (for example, output-proportional mark-ups satisfy ECPR!), usage based models (as in the case of the OFTEL rule under budget balance). It also has been adopted in the absence of price regulation on the competitive segments <sup>(2)</sup>.

Baumol and Sidak (1994a, Section 6) argue that while prices based on information about demand functions have attractive theoretical properties, purely cost based prices are to be preferred for practical purposes. Namely, Baumol and Sidak offer to impose price ceilings for final products equal to the stand-alone costs of these goods. Their theoretical precept was embraced over ten years ago by the US Interstate Commerce Commission to guide the pricing by railroads to shippers <sup>(3)</sup>.

<sup>(1)</sup> This extension is actually suggested by Baumol and Sidak (1994, p. 109), who argue that, in practice, business an entrant gains is not always business lost by the incumbent; entrants offer differentiated services and also 'beat the bushes for customers who were not previous users of the product in question'.

<sup>(2)</sup> New Zealand Telecom is subject only to competition policy on its competitive segments. One of the main objections to ECPR in regulatory proceedings is that, because the TO is guaranteed its margin on the competitive segment even if it loses market share, ECPR could be a 'rationalisation for the continued collection of the monopoly profit despite the introduction of competition' (Tye (1994), Kahn and Taylor (1994)). This fear is indeed a main reason why the New Zealand Court of Appeal initially rejected ECPR.

<sup>(3)</sup> Baumol and Sidak note that a cost-based price ceiling is likely to undermine incentives for productivity or efficiency improvements. To escape this cost-plus feature, they suggest that the 'pertinent stand-alone cost is not the actual cost incurred by the regulated firm, but rather the cost that would be incurred by the entry of a hypothetical entrant'. (We concur with this view on a theoretical ground, but are concerned that it may be difficult to establish a yardstick, especially from a hypothetical entrant.) Two last comments are in order. First, Baumol and Sidak view the stand-alone cost as a price ceiling that the operator is not allowed to exceed. They also recommend setting a price floor equal to marginal (incremental) cost in order to prevent predation by the operator on entrants. Second, stand-alone costs can allow the operator to charge very high prices in the presence of substantial economies of scale or scope.

### 1.5. The optimal access price

In an ideal ‘first best’ world, a necessary condition for optimality would be that the competitors internalise the marginal cost of the bottleneck. So the access price would be set equal to the marginal cost of the bottleneck. Final goods would also obey marginal cost pricing. For example, local calls would cost the consumer this marginal price per unit, and long distance calls supplied by the telephone operator and its competitors would be priced at the marginal costs including this marginal cost of access; the telephone operator’s deficit, would then be covered through a lump sum payment from the State’s budget.

Let us assume that the network operator must balance its budget <sup>(1)</sup>. What are then the optimal prices? The answer was provided long ago by Ramsey (1927) and Boiteux (1956). While these authors considered only final goods, it is clear that their precepts also extend to intermediate goods as well; for, just think of the access good as another final good. Indeed, because competitors have no market power, one can either consider that the network operator produces an access good that is resold without mark-up or envision an equivalent situation in which it would produce the final competitive good itself and would sell it to final consumers. With this transformation the access good becomes a final good produced and the standard Ramsey–Boiteux formulas <sup>(2)</sup> can be directly applied. Relative deviations of prices with respect to marginal costs are inversely proportional to the superelasticities, with a factor of proportionality which is the shadow price of the budget constraint (it must be chosen so that the telephone operator’s budget is balanced). Superelasticities are modified elasticities of demand, which account for possible substitution and complementarity among goods. If we assume that the demands for local calls and long-distance calls are independent, then the superelasticity of local calls is equal to the ordinary elasticity of demand for local calls. In contrast, to the extent that the two long-distance goods are substitutes, their superelasticities are smaller than their respective ordinary elasticities. To see this, consider a decrease in the access price. This decrease cre-

ates more competition for the network operator’s long-distance good and therefore generates a revenue loss for the network operator on the competitive segment, that is a shortfall in the covering of the access deficit. Superelasticities are just meant to reflect the global impact of a change in the price of a good on the network operator’s total profit.

The optimal access price follow from the Ramsey formula directly. It equals the marginal call of access plus a ‘Ramsey term’ which contributes to covering the access deficit.

Note that in this basic model one could equivalently charge an access price equal to marginal cost and levy a tax on the competitors’ output, as long as this tax goes to the network operator. One can thus think of the access deficit contribution as a tax on the competitors’ output. This remark is irrelevant here, but will become important when we introduce the possibility of bypass of the network.

### 1.6. Is ECPR consistent with an optimal access pricing rule?

While we will amend our conclusions on interconnection charges to reflect several realistic departures from the benchmark model, it is worth investigating whether competitors pay the network’s opportunity cost in the benchmark. We analyse three simple cases:

#### *Fully symmetric case*

Suppose that the cost of providing access does not depend on who uses the access, that the network operator and its rivals are equally efficient on the competitive segment, and that the demand functions in this segment also exhibit symmetry. Then the ECPR obtains in this fully symmetric case.

#### *Brand loyalty/name recognition/switching costs*

Suppose next that the telephone operator has captive customers and that demands are linear. The efficient access price is lower than that predicted by the ECPR. It is interesting to understand why this is the case. The telephone operator should charge a higher price than its equally efficient competitors in order to raise revenue from a more inelastic demand. So the network operator’s opportunity cost is high compared to the desired access price, which reflects a higher elasticity of demand on the competitors’ good.

<sup>(1)</sup> Almost equivalently, it could receive a subsidy from the government and tax collection could have a distortionary effect on the economy (see Laffont and Tirole (1994)). More generally, we will make the maintained assumption that the industry as a whole cannot receive government subsidies. Also, the analysis applies to situations where a limited subsidy only partially covers the network’s fixed costs.

<sup>(2)</sup> See, for example Laffont and Tirole (1993, Chapter 3) and the references therein.

### *Learning by doing/network operator's technological superiority*

Last, assume that demand is symmetric (and still linear), but the telephone operator has lower costs on the competitive segment. One can then show that the access price is smaller than ECPR. The intuition for this 'below ECP' rule is again straightforward. The cost differential calls for a lower price for the telephone operator. Yet, under linear demand, there is 'cost absorption'. Conversely, an inefficient telephone operator ought to charge an access price in excess of the ECP level.

## **1.7. A possible implementation of usage-based access prices: the global price cap policy**

### *1.7.1 Partial versus global price caps*

The benchmark model's conclusion that access prices should be usage-based is likely to be challenged. Regulators as well as many economists <sup>(1)</sup> have correctly argued that rules using elasticities and cross-elasticities of demand are hard to implement. Many have therefore called for cost-based rules.

Two arguments are often confused here. First, there is the sheer difficulty of obtaining elasticities <sup>(2)</sup>. The argument is strongest for rapidly changing industries such as telecommunications, and weakest for industries such as electricity where the regulator can obtain relatively precise demand information (although this information need not be as accurate as the firm's. For example, regulators are sometimes suspicious of power companies' requests of the right to give selective discounts, because they cannot verify the utilities' information about their customers' bypass opportunities). Second, besides the technical criticism that elasticities are difficult to estimate, lies the even more important argument that the discretion involved in the determination of these elasticities creates vast opportunities for capture as interest groups gain from influencing the regulator's determination of elasticities and cross-elasticities.

For an industry like telecommunications, the only promising alley for basing the rate structure on demand considerations and avoiding the capture problems is to delegate pricing to the telephone operator under the

form of a price cap. Some will counter that the firm itself may have imperfect knowledge of the demand curve. But imperfect knowledge of the demand curve has never prevented unregulated firms from practising subtle forms of price discrimination, charging low prices for products with elastic demands and high prices for products with low elasticities, adjusting prices to reflect the intensity of competition, and correcting prices upwards when selling substitutes and downwards when selling complements. Indeed, the ideal pricing structure can be obtained by imposing a global price cap <sup>(3)</sup> on the network operator, with the following features.

1. The intermediate good (access) is treated as a final good and is included in the computation of the price cap (this is the definition of a global price cap).
2. Weights used in the computation of the price cap are exogenously determined and are proportional to the forecasted quantities of the associated goods.

As is well known, a price cap induces a firm to select the proper Ramsey structure as definition of the cap and the weights are exogenously fixed at the level of output that will be realised; this result holds for any demand structure and in particular allows for the possibility of strong substitutability between access goods and final goods <sup>(4)</sup>. That is, a global price cap in principle allows a proper usage based pricing structure apparently without a need for the regulator to know the demand functions <sup>(5)</sup>. As we will see, however, the exogeneity of weights is a qualifier to this encouraging result as weights based on realisations of output create some difficulties <sup>(6)</sup>.

<sup>(3)</sup> Global price caps were proposed in Laffont and Tirole (1994, 1996b). We refer to Baumol, Ordover and Willig (1997), Grout (1996), and Schwartz (1995) for careful discussions of the concept.

<sup>(4)</sup> The intuition for this result is straightforward. Let  $\pi(p)$  and  $S^u(p)$  denote the firm's profit and the consumers' net surplus for price vector  $p$ . A social welfare maximising firm subject to a budget constraint would maximize  $\pi(p) + S^u(p)$  subject to  $\pi(p) \geq 0$ . That is, it would maximise  $\pi(p) + \alpha S^u(p)$  for some  $\alpha$  in  $(0,1]$ . When increasing price  $p_i$  by one unit, a profit-maximising firm ignores the impact  $(-q_i)$  on the net consumer surplus, where  $q_i$  is the demand for good  $i$ . On the other hand, a profit-maximising firm subject to price cap  $\sum w_i p_i = w \cdot p \leq \bar{p}$  maximises  $\pi(p) + \beta(\bar{p} - w \cdot p)$  and therefore chooses the proper relative prices if the weights are exogenous and proportional to the realised outputs.

<sup>(5)</sup> We do not discuss here the well-known issue of the firm's incentive to provide service quality under a price cap regulation (see Laffont and Tirole (1993, Chapter 4)).

<sup>(6)</sup> While the following point is somewhat orthogonal to the question at hand, it is worth reminding the reader that there is in practice no pure price cap. Because regulators cannot ignore the firm's past performance, price caps always have some cost-of-service flavour.

<sup>(1)</sup> See, for example Baumol and Sidak (1994) and the WIK-EAC report (1994) in the interconnection context.

<sup>(2)</sup> Note that cost-based rules are also very hard to implement for a regulator. In particular regulators have limited information about incremental costs.



The global price cap assumption is at odds with standard practice. For example, British Telecom's price cap does not include intermediate products. Actually, the very debate about access pricing rules reflects the general view that intermediate and final goods are to be treated asymmetrically. A global price cap denies the specificity of access goods.

It is therefore important to ponder over the logic of a global price cap. Suppose one adopts a partial price cap instead, together with ECPR (this is an approximation of the UK policy until 1997). The network operator then maximises profit subject to this price cap. Suppose further that the weights are chosen proportional to anticipated outputs <sup>(1)</sup>. As we shall see shortly, assuming correct expectations the network operator then biases its rate structure relative to the Ramsey optimum. For example, long distance prices and the access charge are too high while local calls are too cheap. By not including the access charge in the price cap, a partial price cap de facto subsidises non-competitive segments to the detriment of competitive ones.

In contrast, a global price cap, with weights proportional to actual quantities, is able to achieve the Ramsey price structure <sup>(2)</sup>. To achieve the Ramsey structure, weights must be proportional to actual quantities.

#### *Remark on foreclosure and cross-subsidies*

Our discussion has focused on the allocative gain brought about by the price structure induced by a global price cap. Laffont and Tirole (1996a) argues that global price caps offer two additional benefits relative to current regulatory rules or proposals. First, global price caps substantially reduce the firm's incentive to foreclose its rivals through non-price methods (such as delays in interconnection, refusal to unbundle, or costly technical requirements). The possibility of foreclosure has been hotly debated in the context of the RBOC's entry into long distance in the period preceding the February 1996 American telecommunications law, and figures prominently in several disputes in other countries. Intuitively, under a global price cap, the regulated firm can elect to earn most of its income on interconnection and it then has no incentive to reduce its demand for

access. In contrast, current regulations tend to unevenly put more pressure on access charges than on prices of services in competitive segments, and encourage foreclosure. Second, global price caps eliminate another asymmetry in the treatment of product lines that is frequently encountered in existing fragmented regulations. Namely, current regulations provide different incentives for cost reduction or profit enhancement in different product lines. By not balancing incentives properly, they encourage cross subsidies. We refer to our the Laffont and Tirole paper for a more complete discussion.

#### *Remark on profit sharing*

It is important to point out that the phrase 'price cap' has a more general meaning than contemporary regulatory usage. A price cap is logically consistent with profit sharing rules, although its usage has been restricted to situations in which the regulated firm is (theoretically) residual claimant for its profit. That is, once a price cap has been set, any profit sharing mechanism can be superimposed without affecting the implementation of the structure of Ramsey prices.

#### *1.7.2. Global price cap with and without ECPR*

There are two reasons for considering appending ECPR to the global price cap:

#### *Weight-setting*

To implement Ramsey prices through a price cap, weights must be set proportional to the forecasted outputs. A precise forecast may demand information not available to regulators. In practice, weights are often based on recent outputs or revenues (for example, British Telecom's weights are the previous year's revenue shares for the various products). That is, if one does not have good forecasts of actual quantities, one must grope towards them through a *tâtonnement* process. To limit the pricing distortions induced by endogenous weights and to accelerate convergence towards Ramsey prices, regulators must under a partial price cap come up with a reasonable forecast of the total demand on the competitive segment and of the market share of the network operator. While this task is arduous, it is still easier than the one needed to define a partial price cap and consisting in predicting a (demand-, technology-, and regulatory policy-contingent) production for the network operator.

<sup>(1)</sup> In contrast, British Telecom's price cap's weights are proportional to each product's revenue share.

<sup>(2)</sup> See Section 9 of Laffont and Tirole (1994).

### Predation

Regulators and policy advisors are generally concerned by the possibility that incumbents prey on entrants. They usually append to their policies an Areeda–Turner price floor on the incumbent’s prices on the competitive segments <sup>(1)</sup>. That is, the incumbent’s prices should not be set below its long-run incremental costs on the competitive markets.

It is particularly easy for an incumbent to prey under a simple global price cap: it suffices to raise the access charge while lowering the price on the competitive segment so as to satisfy the price cap. Both actions hurt competitors, who may be driven by financial constraints out of the market. They also hurt the incumbent while they are implemented as they ‘mutilate’ its access segment. The harder question is whether the incumbent generally gains from predatory behaviour under a global price cap. But it is clear that there are cases in which predation is profitable. If entry is motivated by the use of yardstick competition to reduce the incumbent’s rent on the competitive markets (see below), eliminating the yardstick enables the incumbent to enjoy higher future rents, so that predation is profitable if entrants are financially weak and can be driven out by a short-lived ‘scissors’ of the kind described above.

By contrast, ECPR protects somewhat entrants from predation by tying access prices and final goods prices. Under ECPR, the incumbent’s only way to prey is actually to charge high prices so as to kill off the market for the competitive good! While not preventing predation, ECPR makes it more costly <sup>(2)</sup>.

On the cost side, ECPR introduces some distortions and also destroys the extreme simplicity of a global price cap by requiring that regulators or courts verify compliance with ECPR. Indeed, ECPR is often criticised as unrealistic <sup>(3)</sup> because it requires measuring the incumbent’s marginal costs — or equivalently margins — on the competitive segments <sup>(4)</sup>.

In practice the ECPR can be used as an upper bound only when real threats of predation exist due to a very asymmetric industrial structure. When competition takes place between two or three strong telecommunications operators, such additional constraints are unwarranted. It should also be noted that the ECPR is one of possibly many anti-predation devices, although a salient one. As for unregulated industries, the profession is still searching for a practical and theoretically satisfactory test of predatory behaviour.

### 1.8. Lack of instruments and multiple goals for interconnection charges

In the idealised case developed above, the setting of the access price had a single purpose; it was only meant to regulate the otherwise undistorted price of the competitors’ good and thereby obtain a proper rate structure on the competitive segment, given that the prices on that segment ought to include a mark-up contributing to the reduction of the access deficit. In practice the need to cover the access deficit is not the only source of distortion.

First, competitors may themselves have market power. This arises in particular when there are large fixed entry costs. Two new distortions may appear in such situations. The competitor(s)’s price includes a mark-up above its perceived marginal cost (that is, its marginal cost plus the unit access charge). And, if entry costs are large, entry may not occur even when it is socially desirable.

Second, the bottleneck segment may not be a pure bottleneck. Although expensive, entry on that segment is feasible. Two prominent paradigms, the bypass and the network duplication paradigms, have been the object of intense policy debate in the telecommunications industry. On the one hand, large business customers (perhaps through a competitive access provider) may be able to bypass the local loop and connect directly to a long distance company in order to economise the mark-up on the access charge. On the other hand, the competitive service providers can install a second network either by themselves or in the case of telecom for example, by entering an agreement with a cable, water or electricity company. Ensuring that the right amount of bypass and of network duplication occurs provides two more targets for public policy.

Third, the network operator may not produce efficiently when the regulators’ imperfect technological informa-

<sup>(1)</sup> See, for example, Baumol and Sidak (1994).

<sup>(2)</sup> The idea of appending ECPR as a predation test to a global price cap is also discussed in Baumol et al. (1997).

<sup>(3)</sup> See, for example OFTEL (1994), Ralph (1994), and WIK-EAC (1994).

<sup>(4)</sup> Compliance with ECPR is however more likely to be checked *ex post*, following a competitor’s complaint, than *ex ante*.

tion prevents them from imposing cost minimisation. A new target for public policy is then to create proper incentives for cost minimisation while not giving up excessive rents to the operator. Should the interconnection policy contribute to the definition of proper incentives for the operator? For example, it has been argued in regulatory proceedings that ECPR neglects the fact that the incumbent's marginal cost on the bottleneck segment may be inflated, which unduly penalises the entrants through a high access price. Baumol's response to this criticism has been that the access price has no role in regulating incentives.

These three classes of further concerns can be analysed as follows. More goals require more instruments. If these further instruments are available to the regulators, what we have said needs only be reinterpreted. However, those instruments may not belong to the regulatory panoply, due in particular to the separation of powers. The access price then needs to contribute to the achievement of conflicting goals. It becomes a 'jack of all trades and master of none'.

#### 1.8.1. Competitors' market power

Let us first consider the case of large fixed entry costs.

##### 1.8.1.1. Profitable entry

Suppose that, due to a large fixed entry cost, the network operator faces a single competitor on the competitive segment. This competitor charges a mark-up above its marginal cost.

The analysis is straightforward when the competitor's profit (net of the fixed entry cost) can be taxed and redistributed to the operator. The profit tax disposes of the issue of capturing the competitor's profit to contribute to the coverage of the access deficit. The access price, as earlier, only guides the competitor's final price. It should be reduced so as to exactly offset the mark-up imposed by the competitor. This yields the same price structure the competitive segment and does not affect the operator's budget relative to the case in which the competitor charges its marginal cost. The new distortion is eliminated by the equal reduction in the access price.

For example, recall from Section 1.6 that in the symmetric case the optimal access price obeys the ECPR. In the presence of competitor market power, the optimal access price is equal to the ECP level minus the com-

petitor's unit mark-up  $m$ . We thus obtain an 'ECP- $m$ ' rule.

The redistribution of the competitor's profit to the network operator is likely to be unrealistic. In the absence of such taxation, the access price is assigned a second role, namely it must substitute for the missing profit tax to limit the competitor's rent. The access price is then raised above the 'ECP- $m$ ' level, and can even in some circumstances exceed the ECP level <sup>(1)</sup>.

##### 1.8.1.2. Unprofitable entry

Our analysis so far has assumed that the competitor is profitable, so that its entry is not an issue. This is not realistic for very large fixed costs; the challenge then becomes to raise the entrant's profit to the break-even point rather than to capture its excess profit <sup>(2)</sup>.

We can apply the previous argument. The 'ECP- $m$ ' rule still obtains (in the symmetric case) when a lump-sum subsidy is feasible. The lump-sum subsidy is then to be included in the computation of the access deficit as yet another unallocated cost.

But suppose now that the lump sum transfer to the entrant is not feasible. Then the access charge must subsidise entry. One correspondingly obtains a 'below ECP- $m$ ' access price. This policy has been extensively used to promote entry in the telecommunications industry. Mercury has benefited in the past from generous access conditions. Similarly the access charge paid by MCI and Sprint has not reflected their high connection cost to the local exchange companies compared with AT&T.

##### 1.8.2. Bypass and network duplication

New policy issues arise when the network operator's natural monopoly position in the 'bottleneck' segment is challenged. In particular, concern has been expressed that bypass and network duplication (PBXs, cellular, competitive access providers), which in specific instances provide cost savings or complement the pub-

<sup>(1)</sup> The exact nature of the correction depends on the kind of competition the firms are waging: see Masmoudi and Prothais (1994) for Cournot competition, and Laffont and Tirole (1994) for price competition.

<sup>(2)</sup> It may be socially beneficial to subsidise an unprofitable entrant because the latter does not internalise the full social value of its entry. However, the subsidy policy may be dangerous for various reasons associated with incomplete information (capture, irreversible errors, etc.).



lic service telephone network, in general may forego substantial economies of scope.

#### 1.8.2.1. Bypass

Large long-distance customers, for example, often can establish a radio link or lay a cable in order to connect directly with the long-distance competitors. Such bypass may be socially efficient. Yet the inclusion of an access deficit contribution into the access price may give large customers an excessive incentive to bypass the local loop.

Regulatory policy must now be concerned with inducing the efficient amount of bypass. In essence, it needs one more instrument. This is where the distinction between access charge and excise tax becomes important. The use of an excise tax to help cover the access deficit frees the access charge from its deficit recovery role. The access charge can then be used to guide large customers in their bypass decisions.

The division of labour between the access charge and the excise tax works roughly as indicated in the following reasonable rule of thumb <sup>(1)</sup>: to provide the large customers with the right signal, the access charge is set near the marginal cost of access. The excise tax then picks up the access deficit contribution.

In practice, however, regulators usually have no mandate to levy taxes in their industry, even when they regulate the network operator's final prices. Assuming zero tax, say, the access price must now arbitrate between two conflicting goals. A low access price prevents wasteful bypass, but substantially increases the operator's deficit. The operator must then set its price on the competitive segment even higher and may well be squeezed out of that segment. This move towards a *de facto* (rather than *de jure*) vertical separation not only deprives the competitive segment from one of its main actors, but also raises the question of how the cost of the local loop and of universal service is to be recovered.

We thus come to the following conclusions. It seems logical to depart from custom and to let the regulator (when feasible) regulate the competitors' prices through

an excise tax affected to covering the access deficit. To be certain, a tax set by regulators is more information demanding than a price decision delegated to the firm. Furthermore, it is risky to extend regulatory powers in this direction. Indeed we conjecture that the combination of limited statutory powers (the regulators are prevented from regulating the entrants) and of a mechanistic access pricing rule is intended to prevent regulatory capture by the industry by freeing entrants from regulatory control.

If one is concerned by this extension of regulatory powers, there is a particularly strong case for allowing substantial quantity discounts in the pricing of access. While quantity discounts are no panacea, they exploit well the fact that high-demand customers are those with the highest incentive to bypass <sup>(2)</sup>. The network operator can profitably offer a menu of two-part tariffs, with the fixed fee inversely related to the per unit charge <sup>(3)</sup>. Small customers would be charged (or rather would select) a low fixed fee with a high per unit charge (which in practice would be collected by the long-distance operators and repaid to the network operator). The main change relative to the present situation is that large customers would choose to pay a high fixed fee in order to face a low per unit charge (close to marginal cost of access).

#### 1.8.2.2. Network duplication

While bypass is, say, the phenomenon in which large customers dispense with the operator's services to reach the competitive service providers, network duplication consists in the competitors' building their own network in order to reach customers who would have not bypassed by themselves. Network duplication substitutes advantageously for individual bypass when bypass connections exhibit returns of scope among customers. Network duplication is often quite costly and its main

<sup>(1)</sup> See Laffont and Tirole (1994) for more details. 'Roughly' refers to the fact that bypass affects the total marginal cost of long-distance calls, and to the regulator's desire to extract some of the bypassers' rent.

<sup>(2)</sup> On this, see Laffont and Tirole (1990) and Curien, Jullien and Rey (1993).

<sup>(3)</sup> It is interesting to note in this respect that the California Public Utility Commission accepts the argument according to which quantity discounts on access prices can prevent uneconomic bypass, and estimates that the corresponding efficiency gain overshadows a potential favoritism in favour of AT&T and to the disadvantage of smaller interexchange carriers (PUC of the State of California (1994, pp. 123–126)). The California decision came in response to GTEC's 'Switch Access Volume Election' (SAVE) proposal. The SAVE plan offers volume discounts on access to compete with high-speed digital private lines. (Interestingly, a higher threshold of originating minutes of use than terminating minutes of use is required to qualify for SAVE credits. Bypass is more common at the originating end than at the terminating end. This fact also provides a rationale for charging higher access prices at the terminating end than at the originating end.)

justification is that it facilitates the regulatory task of providing incentives for low costs in the local loops.

Taking for granted the optimality of network duplication, the arguments laid out in the discussion of entry in the competitive segment resurface in the analysis of entry in the complementary segment. One could either encourage the building of a rival network through a subsidy, or, if subsidies are not available, charge high access prices in order to discourage the use of the incumbent network by its competitors <sup>(1)</sup>. One should also note that, if network duplication is contemplated, the rival networks should face the same universal service obligations as the incumbent network. That is, networks should be treated symmetrically.

### 1.9. Access charges and incentives for cost minimisation

The analysis has until now ignored the important issue of the network operator's incentive to minimise its costs. One may wonder, as some policy analysts and judiciary do, whether access charges should also be employed for cost-minimisation purposes.

#### 1.9.1. Incentives to minimise bottleneck cost

Laffont and Tirole (1994) offer a comprehensive analysis of the impact of incentive considerations on pricing and shows that, under some conditions, pricing and incentives issues are decoupled <sup>(2)</sup>. That is, the price formulas are the same as when the regulator has perfect information about the technology and incentives are not an issue; they still hold for the realisation of marginal costs, whether or not these marginal costs are inflated by poor incentives. This decoupling may be a reasonable rule of thumb in the absence of detailed information about the cost function. (Incidentally, it provides a foundation for Baumol's position that the access price is not meant to correct poor incentives for cost minimisation.)

<sup>(1)</sup> Let us here note that the latter policy (high access charge as an inducement for network building) may require discrimination in access prices; for, most suppliers in the competitive segments using the local loop will not build their own network and will need to face a reasonable access price.

<sup>(2)</sup> This is the 'dichotomy property'. Technically, it holds whenever quantity levels do not affect rates at which technological improvements can be converted into rents by the firm. Incentives are then provided by the cost reimbursement rule.

#### 1.9.2. Incentives to minimise cost on the competitive segment

The previous decoupling applies equally well to the operator's incentives on the competitive segment as long as the behaviour of its competitors does not provide information about its operator's cost structure on that segment. That may be a reasonable assumption if the competitors use a different technology. In contrast, when technologies on the competitive segment are similar one may use the competitors' performance as a benchmark to control the operator's efficiency. This 'yardstick competition' is indeed one of the main arguments for the creation of competition, although it has been left out of the debate on access pricing.

A new insight is that entry may be desirable even if the entrants face a fixed entry cost and produce a close substitute on the competitive segment. The gains from yardstick competition are larger, the less regulators are informed about the incumbent technology on the competitive segment, and the more similar the incumbent's and the competitors' technologies <sup>(3)</sup>. If there is a fixed cost which renders entry unprofitable then the only way to induce entry is by lowering the access price further, and it may be worth doing so even if the variety advantage brought by the entrant is small, because the entrant now brings a second advantage, namely, the yardstick.

### Conclusion

Let us summarise our main insights, and mention a few unexplored questions. We developed a conceptual framework in which competition is motivated by product differentiation, cost differentials and/or (in the last section) yardstick considerations. This framework sheds light on the various policy proposals and suggests improvements.

1. A discussion of an access rule without reference to the rest of the regulatory environment has limited interest. The quality of an access pricing rule depends on the determination of prices for the final products.
2. If the regulators are not constrained in their panoply of instruments the access charge is deter-

<sup>(3)</sup> We should also note that in practice yardstick competition is likely to reveal more information about the incumbent's technology only if the competitors' market share is nonnegligible, so that the incumbent and its rival(s) operate in similar parts of the cost function.

mined solely with the concern of avoiding inefficient creamskimming and participating to the recovery of the access deficit. The 'generalised access charge' (access charge, plus excise tax on the competitors' goods) depends on the use that is made of the access. This conclusion ought to be born in mind when assessing the rule of non-discrimination in access contained in the European Open Network Provision. It questions the general policy of the Commission against demand-based access pricing. The cost based argument seems to come from a direct application of the foreclosure theory for competitive policy to regulated environments. But, regulatory rules can change the private and social costs and benefits of foreclosure, and in our view further analysis is required to justify the cost based doctrine. (Actually as we explain in Part three, some steps in the direction of demand-based access charges in telecommunications are taken in a recent document.)

We found that ECPR is optimal under symmetry assumptions about cost and demand functions. ECPR must be amended to account for asymmetries between the integrated firm and its competitors. The access price should lie below its ECP level if the integrated firm has captive customers or is technologically more efficient.

3. It is well known that ECPR, like any other access pricing rule, does not supply any recommendation as to the integrated firm's pricing of competitive products. There are two standard stances on this issue.

One practically-oriented view is to base the pricing of the integrated firm's goods purely on costs, as under the traditional fully-distributed-cost (FDC) methodology. Basing prices on cost may have perverse incentive implications. Further, fully distributed-cost pricing induces cherry-picking by the competitors, and thereby generates allocative inefficiencies and jeopardises the recovery of the access deficit. It also creates inefficient bypass.

The Ramsey approach, which limits incentives for cherry-picking, is theoretically more satisfactory. The main difficulty with this approach is that the regulators are unlikely to hold the required information about demand. In our view, the main hope for obtaining Ramsey-orientated prices is the delegation of pricing to the integrated firm. In this respect we

have discussed 'BT-style' partial price caps, which apply only to final goods, and global price caps, which also include intermediate goods. A partial price cap has two flaws compared to a global one <sup>(1)</sup>. First, the omission of the intermediate goods in the integrated firm's price cap subsidises price increases on the competitive segments and biases the rate structure toward high prices in those segments and low prices in the non-competitive ones. Second, this omission complicates the setting of the weight of competitive prices in the price cap, as it requires forecasts of the integrated firm's market share besides the prediction of total demand in the competitive segment. A global price cap penalises increases in both access prices and final prices and induces the integrated firm to price discriminate very much the way an unregulated firm would do, except that its entire price structure is brought down by the cap.

4. Still assuming that the regulators have enough instruments to free the access price from the fulfilment of multiple conflicting goals, these principles can be reinterpreted in a straightforward manner when fixed entry costs in the competitive and bottleneck segments are brought into the picture. For example, when the competitors have market power and impose a mark-up and under demand-and cost-symmetry, ECPR becomes an 'ECP-*m*' rule as long as the competitor(s)' profit (loss) can be taxed (subsidised) and contribute to the integrated firm's budget. Or, quite importantly, bypass can be efficiently regulated by choosing an access price close to marginal cost and by using an excise tax on the competitors' products to recover the access deficit as long as regulators can levy such a tax and pay the proceeds to the integrated firm.
5. The separation of powers has deprived regulators from many useful instruments such as excise taxes on competitors' products, profit redistribution, or lump-sum entry subsidies. Consequently, regulators are forced to use the access price to arbitrate among conflicting goals (recover the access deficit, prevent inefficient bypass and network duplication, capture competitors' rents or ensure that competition is viable, etc.). We have described the direction of the corrections that must be made in the access rule to reflect the lack of instruments.

<sup>(1)</sup> Laffont and Tirole (1996a) discuss the advantage of global price caps from the point of view of exclusionary practices.

6. Decoupling access rules and incentive issues seems a reasonable rule of thumb in our state of knowledge.

## 2. Bottleneck access and market foreclosure in an unregulated environment

### Introduction

We now consider situations where an unregulated monopoly operates a facility that is an essential input to competitive service providers. Such situations raise market foreclosure issues, and have given rise to the so-called essential facility doctrine. We first present those two notions, and then briefly survey the recent literature and its policy implications.

#### 2.1. What is foreclosure?

According to the received definition, foreclosure refers to any dominant firm's practice that denies proper access to an essential input it produces to some users of this input, with the intent of extending monopoly power from one segment of the market (the bottleneck segment) to the other (the potentially competitive segment). The excluded firms on the competitive segment are then said to be 'squeezed' or to be suffering a secondary line injury. Essentiality means that the dominant firm's product cannot cheaply be duplicated by users who are denied access to it, a feature which is common to many networks. Examples of essential facilities or bottlenecks to which competition law has been applied include a stadium, a railroad bridge or station, a harbour, a power transmission or a local telecommunications network, and a computer reservation system <sup>(1)</sup>.

The foreclosure or essential facility doctrine states that the owner of such an essential facility has an incentive to monopolise complementary or downstream segments as well. This doctrine was first discussed in the USA in *Terminal Railroad Association v U.S.* (1912), in which a set of railroads formed a joint venture owning a key bridge across the Mississippi River and the approaches and terminal in Saint Louis and excluded non-member competitors. The Supreme Court ruled that this practice was a violation of the Sherman Act. A version of the doctrine was invoked by the European Court of Justice

in the celebrated *United Brands* (1978) decision, in which it held that United Brands Corporation enjoyed substantial market power in the banana market in Europe and engaged in exclusionary practices in related markets (distribution, ripening) <sup>(2)</sup>. The *Clear* case in New Zealand (see Subsection 1 above) provides an example of application of the doctrine to telecommunications, where the dominant operator (Telecom)'s network has been considered as an essential facility.

Foreclosure can take several forms. It can be complete, as in the case of a refusal to deal or of an extravagant price for access to the essential facility ('constructive refusal'), or partial, as when the bottleneck owner favours some downstream firms (perhaps, its subsidiary) to the detriment of other competitors who still have (limited) access to the essential facility. It can also be performed in various ways.

- (a) The bottleneck owner can integrate vertically with one or several firms in the complementary segment. For example, computer reservations systems were developed by major airlines. Before the Civil Aeronautics Board (CAB)'s 1984 famous decision, it was perceived that smaller airlines, especially those competing head to head with the integrated firms, had to pay a high price for access to the reservation systems and received poor display of their flights on the travel agent's screen (a key competitive disadvantage given that most travel agents do not browse much through screen displays). The CAB attempted to impose equal access in price and quality to what are perceived to be essential facilities, namely computer reservation systems. (Similarly in 1988, the European Commission imposed a fine on Sabena for denying access to its computer reservation system to the price-cutting airline London European.) Whereas the CAB did not call for the major airlines' divestiture of their computer reservation systems, in the same year American courts forced AT&T to divest its regional operating companies (known as the RBOCs). Other examples of forced vertical separation include the UK brewing industry, in which, following an investigation by the Monopoly and Mergers Commission in 1989, the 'majors' were instructed to divest pubs, an essen-

<sup>(1)</sup> Extensive legal discussions of foreclosure can be found in Areeda (1981) and, especially, Hancher (1995).

<sup>(2)</sup> The most recent European case involves the Port of Roscoff, accused of favouring one ferry operator and foreclosing others.

tial facility <sup>(1)</sup>, and the British rail system, in which restructuring creates a separate provider of access, Railtrack.

- (b) The integrated firm can refuse to deal with potential competitors. Relatedly, it may engage in tie-ins and refuse to unbundle, thereby denying access to the essential facility. For example, in *Port of Genoa* (1991), the European Court of Justice held that the harbour is an essential facility and that its use should not be reserved to the undertaking managing it <sup>(2)</sup>. A number of cases involve the requirement by a durable good manufacturer with market power that repairs, maintenance or spare parts be provided by the manufacturer <sup>(3)</sup>.
- (c) In the presence of economies of scope or scale generated by the cooperation between firms in the same market, a dominant group of firms may put its competitors at a disadvantage by refusing to cooperate. Famous cases include *Aspen Skiing Co. v Aspen Highlands Skiing Co* (1985), in which the common owners of three mountains on the site first offered a low percentage and then discontinued the All-Aspen ski passes which enabled skiers to use these mountains as well a fourth one independently owned; and *Associated Press v United States* (1945), in which members of the newspapers cooperative could block membership by competing newspapers. Such cases have obvious implications for network industries <sup>(4)</sup>.
- (d) Short of integration, the bottleneck owner can grant exclusivity to a subset of firms on the complementary segment, and thus de facto exclude their rivals. For example, the European Commission has investigated the 65-year contract

between Eurotunnel on the one side, and British Rail and SNCF on the other side, allocating the entire capacity to the latter two companies. In New Zealand, the Court held that the exclusive rights granted to Avis and Hertz for the operation from the Auckland airport terminal building by Auckland Regional Authority violated sections 27 and 36 of the Commerce Act.

- (e) Another instrument in the 'forecloser's' toolbox is second- and third-degree price discrimination. Third-degree discrimination consists in charging different (cost-adjusted) prices to different customers (e.g. special fares for students). It generalises exclusivity arrangements by favouring some customers over the others, but gives the bottleneck owner some flexibility in serving discriminated-against customers. Even if outright third-degree price discrimination is prohibited, the bottleneck owner may be able to duplicate it in an apparently anonymous way, that is through second-degree price discrimination. For example, a loyalty program offered to all or rebates based on the rate of growth of purchases may target specific customers even though they formally are available to all customers: in New Zealand, the new contracting agreement between Clear and Telecom, which applies to all 'large enough' operators, is a recent example. Similarly, substantial price discounts may allow the survival of only a few customers; for instance, a large enough fixed (that is, consumption independent) fee transforms a potentially competitive downstream industry into a natural monopoly industry. Such considerations (besides many others) played a role in the process of enacting the Robinson-Patman Act in the USA in 1936 <sup>(5)</sup>. There was in particular a concern that independent wholesalers or retailers might not be able to compete with powerful chains buying their supplies at favourable prices.

<sup>(1)</sup> Snyder (1994) performs an event study analysis of this industry and provides some evidence of non-competitive behaviour. See also Slade (1995).

<sup>(2)</sup> A related case is the *Sealink* decision (1992), where the same company operated ferry services and controlled the harbour.

<sup>(3)</sup> See, for example, in Europe, *Hugin v Commission* (1979), in which a manufacturer refused to supply spare parts for its cash machines and the Commission held that the manufacturer had a dominant position on its own spare parts. A recent and hotly debated case in the USA is *Kodak*, who refused to sell replacement parts for photocopiers to owners unless the latter agreed not to use independent service organisations (see Borenstein et al (1995) and Shapiro (1995) for a discussion of this case).

<sup>(4)</sup> For example, *Otter Tail Power Co v United States* (1973) established a (controversial) duty for a vertically integrated power company to supply other companies. In *Aer Lingus* (1992), the European Commission condemned Aer Lingus for refusing to interline (a technique enabling the marketing of single tickets for combined flights) with British Midland.

## 2.2. Envisioned remedies

A number of remedies have been considered by competition law practitioners, and it may be useful to review here the most prominent ones. It is convenient to group those existing policies into five categories:

<sup>(5)</sup> Interestingly, in *Hoffman La Roche*, the European Court upheld the Commission's condemnation of purchasing agreements or loyalty rebates while asserting the company's right to offer volume discounts as long as they are extended to all customers.



### *Structural policies*

Structural policies such as divestitures and line of business restrictions are often considered in last resort, as they may involve substantial transaction costs of disentangling activities and may jeopardise the benefits of integration. Although compete separation is imposed in specific instances (as for the AT&T 1984 divestiture), milder forms of vertical separation are sometimes considered; for instance, the essential facility may be commonly owned by all users, with the provision that new entrants be able to purchase shares and membership into the network ‘at a reasonable price’ (as in the *Associated Press* case mentioned above). The joint ownership of an essential facility by competitors must then be granted an exemption from certain antitrust provisions (as is done for example for certain types of R & D joint ventures).

### *Access price control*

In the tradition of fully distributed cost regulation of access in regulated industries, antitrust authorities sometimes compare the price of access with some measure of its cost. The principle of such a comparison was for example accepted by the European Court in *United Brands* (1978), although it did not apply it in the specific instance. As is well known, the measurement of marginal cost is a difficult empirical matter, while the allocation of common costs among product lines has weak theoretical underpinnings. Clearly, the antitrust authorities lack the staff to conduct extensive cost studies; at best can one put the onus of proving overpricing on the excluded competitors, who may well have better cost information than the authorities.

### *Access quantity control.*

Instead of trying to define a ‘right’ access price, the authorities sometimes focus on the quantity of access. For example, following an investigation of the Eurotunnel 65-year exclusivity contract, allocating the entire capacity to British Rail and SNCF, the European Commission asked that 25 % of each operator’s capacity be allocated to new entrants for passenger and freight services.

### *Price linkages*

Antitrust authorities often try to use other prices — for access or retail goods — as benchmarks for the access price. A famous rule, discussed in Section 2.1 and variously called the efficient component pricing rule

(ECPR), the Baumol–Willig rule, the imputation rule, the parity principle, and (perhaps confusingly) the non-discrimination rule, links the integrated monopolist’s access and retail prices. Namely, the access price charged to competitors should not exceed the price charged by the integrated firm on the competitive segment minus the incremental cost of that firm on the competitive segment. For example, the ICC has expressed a preference for the use of ECPR in railroad disputes in the USA. There are also various forms of mandated linkages between access charges. The bottleneck firm may be forced to offer the same tariffs to all users, or even to charge a single per unit price. Or, it may be required to charge a price of access not exceeding the price charged for final use of the bottleneck segment (for example, the access charge for the local telephone network may not be allowed to exceed the price of local calls for residential or business consumers). Last, there may be mandated linkages between several firms’ access prices, as in the case of reciprocity in access charges for two competing telecommunications networks (each network being an essential facility for the other).

### *‘Common carrier’ policies*

This policy amounts to turning the vertical structure of the industry upside down. In a complementary goods industry, labelling one segment the ‘upstream segment’ and the other the ‘downstream segment’ may appear to be purely semantic. However it is not, since the downstream firms not only purchase goods (inputs) from the complementary segment but also are the ones who interact with the final consumers. The relevance of this question is illustrated (in a regulatory context) by Order 436 which created a structure that allows US gas producers to directly sign contracts with the gas customers (and purchase access from the pipelines bottleneck) rather than staying mere suppliers of inputs to pipelines packaging a bundle of production and transport to final customers.

### *Disclosure requirements*

Another tool in the policymaker’s box is the requirement that contracts for intermediate goods be made public, with the hope that more ‘transparency’ in supply contracts will promote downstream competition. Note that transparency is not equivalent to the prohibition of access price discrimination among buyers. A disclosure requirement does not preclude different tariffs for different buyers.

### 2.3. The theory of market foreclosure

For all its prominence in competition law, the notion of foreclosure until recently had poor intellectual foundations. Indeed, the intellectual impetus in the late seventies (reflected in the American antitrust practice of the 1980s) cast serious doubt about its validity. In particular, the Chicago School <sup>(1)</sup> thought that the whole concept resulted from a confusion about the exercise of monopoly power. It argued that a bottleneck monopolist could earn monopoly profit on the corresponding segment, but could not extend its market power to related segments; for example, in the absence of efficiency gains, vertical integration could not increase the profitability of the merging firms. Relatedly, it questioned the rationale for excluding downstream customers, who could be the source of extra monopoly profits. The Chicago School view has had the beneficial effect of forcing industrial economists to reconsider the foreclosure argument and to put it on firmer ground.

Consider the following quintessential bottleneck situation: an upstream monopolist, the bottleneck owner, produces a key input for downstream use. There is potential competition in the downstream segment, but it can develop only if competitors have proper access to the essential input. The foreclosure doctrine states that in this situation the bottleneck owner has an incentive to restrict or deny access to the intermediate product to some or most of its potential buyers, and thereby to favour a downstream independent firm or a downstream affiliate. This doctrine maintains that foreclosure aims at extending the bottleneck's monopoly power to the downstream segment. The thrust of the Chicago School critique of this doctrine is that there is only one final product market and therefore only one monopoly power to be exploited, and that it is not obvious how the upstream monopolist could further extend its monopoly power. The reconciliation of the foreclosure doctrine and the Chicago School is based on the observation that an upstream monopolist in general cannot fully exert its monopoly power without engaging in exclusionary practices <sup>(2)</sup>. This fact is little acknowledged except in the specific contexts of patent licensing and of franchising. A patent-holder is the owner of an essential facility, namely a technology that can be used as an input in pro-

ductive processes. The patent-holder is unlikely to make much money if it cannot commit not to flood the market with licenses; for, if everyone holds a license, intense downstream competition destroys the profit created by the upstream monopoly position. Therefore, a patent-holder would like to promise that the number of licenses is limited. There is however a commitment problem: once the patent-holder has granted  $n$  licenses, it is then tempted to sell further licenses. It thereby depreciates the value of the existing  $n$  licenses. Such expropriation is *ex post* profitable for the licensor, but reduces its *ex ante* profit. A similar point can be made for franchising. Franchisees are unlikely to pay much to franchisers if they do not have the guarantee that competitors will not set shop at their doorsteps.

The licensing and franchising examples involve binary decisions for input transfer (grant or not a license or franchising agreement). But the commitment problem is very general and extends to situations in which downstream firms purchase variable amounts of the essential input. It is then not surprising that the loss of monopoly power associated with the commitment problem is more severe, the more competitive the downstream segment. This proposition has two facets. First, the upstream bottleneck's profit is smaller, the larger the number of downstream firms. Second, for a given number of downstream firms, the upstream profit is smaller, the tougher is downstream competition.

Bottlenecks are rarely pure bottlenecks. They most often compete with inferior goods or services. In the presence of such bypass opportunities, an upstream bottleneck owner must face both the commitment problem and the threat of second sourcing by the downstream firms. A couple of interesting insights result from this extension of the basic framework. First, a vertically integrated firm controlling the bottleneck in general may want to supply a limited but positive amount of the essential input to the downstream affiliate's competitors, who would otherwise purchase the inferior good. The prospect of productive inefficiency creates scope for profitable external sales by the bottleneck owner. Second, and relatedly, bypass possibilities create a distinction between two ways of restoring monopoly power, vertical integration and exclusive dealing. While exclusive dealing does not enable the bottleneck owner to supply several downstream firms, vertical integration in contrast provides enough flexibility to supply non-affiliates and yet favour the affiliate.

<sup>(1)</sup> See, for example, Bork (1978) and Posner (1976).

<sup>(2)</sup> See Hart and Tirole (1990), McAfee and Schwartz (1994), O'Brien and Schaffer (1992).



#### 2.4. Policy and business strategies implications

This analysis has three broad policy implications. First, it does matter whether the more competitive of two complementary segments lies upstream or downstream: prices are always lower when the bottleneck owner lies upstream. This result is robust to the nature of competition in the complementary segment, to the existence of bypass opportunities, and to the vertical structure of the industry (independent or vertically integrated bottleneck). Intuitively, an upstream bottleneck location has two benefits from a social welfare point of view. First, it creates a commitment problem not encountered by a downstream monopolist and thus reduces monopoly power. Second, in the presence of bypass opportunities, an upstream location of the bottleneck prevents productive inefficiency by creating a stage of competition that eliminates inferior substitutes. The analysis thus supports common carrier policies.

The second policy implication is that non-discrimination laws may have the perverse effect of restoring the monopoly power that they are supposed to fight. When an upstream bottleneck practices foreclosure by discriminating among competitors, it is tempting to impose a requirement that all competitors be offered the same commercial conditions. Non-discrimination rules however benefit the upstream bottleneck because, by forcing it to sell further units at the same high price as the initial ones, they help the bottleneck commit not to flood the market. A non-discrimination law is thus a misguided policy in this situation.

The third policy implication is that the efficient component pricing rule, which was designed for a regulated environment, but is also used in antitrust contexts, often has little bite in the unregulated environment. As mentioned, the ECPR states that the transfer price to be paid by competitors for access to the bottleneck should not exceed the difference between the bottleneck owner's price of the final good and its incremental cost on the competitive segment. As pointed out by William Baumol in testimonies, ECPR only provides a link between access and final prices and is therefore only a partial rule. Moreover, the higher the final price, the higher the access price can be. In an unregulated environment, an integrated firm with upstream market power can thus exercise its market power by setting a high price for the final good and, at the same time, set a high access charge to prevent other firms in the competitive segment from becoming effective competitors.

This analysis has also implications for business strategy. Interestingly, while the desire to foreclose in general motivates vertical integration, it may alternatively call for divestiture. For example, a rationale for the recent and voluntary divestiture of AT&T's manufacturing arm can be derived from the foreclosure doctrine. With the impending competition in telecommunications between AT&T and the RBOCs, the latter, who are major buyers of AT&T equipment, would have been concerned that the AT&T manufacturing arm would exclude them in order to favour its telecommunication affiliate. The RBOCs might therefore have turned to alternative manufacturers. The 'smaller-customer-base' effect of vertical integration may dominate the foreclosure effect, and thus divestiture may be preferred to vertical integration (this rationale is actually related to the official justification for this divestiture).

Also, Chemla (1995) develops the (Williamsonian) argument that downstream competition protects the bottleneck's investment against expropriation in a situation in which the downstream firms have non-negligible bargaining power. There is then a general trade-off between foreclosing competition downstream so as to exploit monopoly power and preserving competition there in order to protect upstream rents.

#### 2.5. Potential defences for exclusionary behaviours

Vertical integration and foreclosure may also have social merit in some instances. For example, unrestrained competition may sometimes lead to excessive entry and duplication of fixed costs, and vertical foreclosure may help reducing this excessive entry. Also, vertical integration may help the upstream and downstream firms to achieve a better coordination, for example by providing better incentives to monitor firms' efforts; foreclosure then is an undesired by-product of a useful institution. We briefly mention here these possible defences.

##### *Excessive entry in the downstream competitive segment*

Suppose for example that there is a large number of potential firms for the production of the downstream good, and that each downstream firm chooses has to incur a fixed, irrecoverable cost to enter the market. If all downstream firms produce the same homogenous good, efficiency considerations would dictate to have only one downstream entry. There is then technologically excess entry whenever more than one downstream firm is active in equilibrium. But entry enhances down-

stream competition and there is therefore a trade off between technological efficiency and competitive pressure. Foreclosure by the upstream monopolist de facto limits the level of entry. The question is then whether this private control of entry goes in the direction of the socially optimal level of entry.

In the absence of vertical integration, and assuming imperfect oligopolistic competition (e.g. Cournot competition), there typically is more than one active downstream firm: those firms have an incentive to enter as long as the expected profit remains higher than the fixed cost of entry and, because of the Coase problem, the upstream bottleneck owner has an incentive to go along: as long as the entry a new firm can generate profits — even if it adversely affects the profits of the downstream firms already present in the market — the upstream monopolist has an incentive to let this new firm enter and get at least some of its profit.

Under vertical integration, the bottleneck owner forecloses the downstream market. As a result, the number of active downstream firms is the efficient one, but the price is the monopoly one. Hence if the duplication of the fixed cost is particularly harmful, vertical integration may yield a socially better outcome than no integration<sup>(1)</sup>. Note however that the validity of this argument may be difficult to assess in practice, since the characterisation of the socially optimal number of firms is generally a complex matter.

#### *Forbearance as a reward to innovation*

The monopoly obtained through foreclosure activities may in fact help compensate, or even be necessary to compensate the bottleneck for its investment or innovative activity. For example, one might imagine that no prospective licensee would want to pay for the use of a new technology if she knew that the licensor could ‘flood the market’ with similar licensees. A similar

argument can be made for franchises. The efficiency defence is here identical to that underlying the patent system. In both cases society is willing to tolerate static inefficiency, such as monopoly pricing, in order to promote dynamic efficiency. So, the general issue is: to what extent is forbearance an optimal mechanism for providing innovators with a rent? The reward-to-innovation efficiency defence provides a key to the analysis of when antitrust authorities may want to force access to a bottleneck. It would not be serious to mandate competitors’ access to each and every aspect of a firm’s activity on an unbundled basis. As illustrated in the *Microsoft* case as well as in the telecommunications industry, one must be careful in defining which bundles competitors are entitled to have access to (all the more as Microsoft relentlessly expands the definition of an operating system and that telecommunications networks and products evolve rapidly). Furthermore, as recognised in *Aspen*, one cannot impose a general duty to deal with competitors. One plausible dividing line to answer the question of when it is most desirable to force access is the following: is the origin of the bottleneck increasing returns to scale or scope (as in the case of a bridge, a stadium, or a news agency) or an historical accident? Or does the bottleneck result from a previous innovative strategy? Intervention to avoid foreclosure and consequently to reduce the bottleneck profit seems more warranted in the former than in the latter case.

#### *Monitoring benefits of vertical integration*

Benefits of vertical integration are often mentioned as efficiency defences. For example, control of a supplier by one of the buyers may put someone in charge of making sure that the technological choices of the supplier are in the best interest of the buyers. To be certain, the integrated buyer may then use its control right over the supplier to engage in non-price foreclosure, for instance by insisting on technological specifications that are biased in its favour. And it may overcharge the buyers while keeping an internal transfer price equal to marginal cost and thus practice price foreclosure. These foreclosure practices are then seen as an undesirable by-product of an otherwise desirable activity, namely monitoring.

#### *Costly divestitures*

Forcing vertical separation may involve a disruptive cost of disentangling deeply intertwined activities. Hence, even if they would have prohibited the merger

<sup>(1)</sup> See Rey and Tirole (1996) for such an example, based on the standard Cournot model with linear cost and demand. It is then shown that foreclosure can be socially desirable when the ratio between the monopoly profit and the fixed cost of entry lies between 1/2 and 3/2. See also Vickers (1995) for a related analysis of the relative cost and benefits of vertical integration in the context of a regulated upstream monopolist. There again, vertical integration leads to a higher (regulated) access price (it is more difficult to extract the information from the integrated firm, hence the incentive scheme must be more high-powered, resulting in a higher access charge) but less duplication of the fixed cost (because of foreclosure). Vickers’ model is staged in a regulatory context in which (i) the regulator controls the firm’s price but not profit, (ii) the regulator operates direct transfers to the firm and (iii) the regulator has no statutory power to regulate entry in the related market.

of two vertically-related firms, antitrust authorities may not force the firms to divest when faced with the fait accompli of vertical integration.

*Costly expansion of capacity or change of standards in order to provide access*

In practice, a vertically integrated firm may face a higher cost when supplying competitors than for internal transactions, either because the competitors are new entrants and upstream decreasing returns to scale make marginal units more costly to supply than inframarginal ones, or because there is a genuine asymmetry between the costs of supplying the downstream affiliate and its competitors. In essence, this efficiency defence amounts to saying that there is no foreclosure because discrimination among competitors is cost based.

*Fear of being associated with inferior downstream partners who might hurt the firm's reputation*

There may be negative externalities on the upstream firm that are not socially desirable. For example, misbehaviour by a downstream firm may spoil the reputation of other downstream firms and of the upstream bottleneck. This argument, which relies on the existence of monitoring of the downstream firms, is often invoked for example in a franchising context, and used to justify strict quality controls.

*Concern about the downstream firms' credit worthiness*

In some cases, the upstream firm may be concerned about incurring a trade credit risk and may legitimately refuse to supply on credit a buyer that is on the brink of bankruptcy.

*Free-riding by the downstream units on the marketing expenses of the upstream firm*

This argument states that the upstream firm must be able to recoup marketing expenses that will benefit downstream units. It is related to the above discussion of Chemla's work and also to the argument of forbearance as a reward to investment.

*Cream-skimming and other Ramsey arguments linked with the recovery of joint costs*

Foreclosure may enable the upstream bottleneck to recoup its investment or fixed cost. This argument has already mentioned above, when we discussed whether overall profit of the upstream bottleneck is sufficient to

compensate the bottleneck for its investment. But even if the overall profit offsets the investment cost, one must wonder whether the structure of profits is efficient. Suppose for example that the bottleneck serves two unrelated downstream markets, the elasticity of demand being higher in the first one <sup>(1)</sup>. Both profitability and social welfare considerations then dictate that the final price be smaller in the market with high demand elasticity; hence upstream investment costs should optimally be recouped by charging more the market with low demand-elasticity. Suppose now that the high demand-elasticity market is served by a downstream monopolist or duopoly while the other market has a large number of downstream firms. In the absence of foreclosure, a high price would be charged in the first market while, because of the Coase problem, consumers would pay the industry marginal cost in the second market. The recovery of the upstream investment cost would thus have the wrong structure of relative prices in the two markets. A better structure would be obtained, albeit at the cost of increased monopoly power, by allowing foreclosure in the market with low-elasticity, while forbidding it in the other market.

*Universal service*

It is sometimes argued that universal service obligations imposed by the regulator or the law should be compensated by a greater leniency vis-à-vis foreclosing behaviours (see, e.g., the *Corbeau* decision in Europe). This argument is a variant of the more general argument that fixed costs must be recouped by market power in some market. Again one must then question whether foreclosure is the most efficient means of creating market power. We refer the reader to Part two of this report for a discussion of the provision of universal service in a competitive environment.

### 3. Two examples: transportation in the gas and the electricity industries

In the electricity industry, recent technological changes now allow to generate electricity efficiently at a small scale. This opens the possibility to introduce competition at the generation stage, and the main bottleneck seems now to be at the transmission and the distribution

<sup>(1)</sup> The following argument builds on considerations that usually arise in a regulatory context (in particular in the telecommunications industry), but are here transposed and adapted to a deregulated environment.

stages. Reforms in the USA and the UK have tried to organise the competition between generators, through the regulation of the transmission network and by allowing generators and (large) buyers to directly contract with each other.

In the UK the new system was decided by the Electricity Act in July 1989 and came to effect in Britain, Wales and Scotland into 1 April 1990. (In Northern-Ireland the Electricity Order came into order in 1992.) In Britain and Wales, the generating plants have been divided between three new companies: Nuclear Electricity took possession of all nuclear plants and has remained in the public sector while all the non-nuclear plants were divided between two private companies, National Power and PowerGen <sup>(1)</sup>. The twelve Area Boards responsible for distribution were privatised in November 1990 and converted into Regional Electricity Companies (RECs), who have partial local monopoly and moreover jointly own the National Grid Company (NGC), where all transmission assets have been vested <sup>(2)</sup>.

The competition between the suppliers is organised through the pool. Under the tutelage of NGG, the pool provides many of the functions provided by a control area in the United States. Besides maintaining the technical characteristics of the network (frequency, voltage, stability, etc.) the pool dispatches generation to match supply and demand in real time. It operates as a 'day-ahead' spot market: everyday by 10 a.m. each generator submits a schedule detailing for each plant available the price at which it is willing to supply power on the following day for each half-hour period. The pool administrator, considering the schedules and the demand forecasts, determines a merit order for the plants and derives the marginal price, which, together with an element of capacity, determines the Pool Input Price (PIP). The PIP plus the transmission price constitutes the Pool Output Price (POP) which is paid by all buyers.

At the time of the privatisation, the production prices were not regulated, the pool system being thought to

ensure competition. However, despite the capacity surplus on the British market, the operation of the pool has not caused prices to be driven down by competition and resulting efficiency gains. The two large companies, National Power and PowerGen have been suspected of exploiting their market power to maintain prices at an artificial high level, as pool prices have been risen significantly since 1990. The regulator (OFFER) agreed not to refer the two companies to the Merger and Monopoly Commission if they consented to sell some of their plants to potential competitors and to adhere to a price cap. The price cap on the production prices became effective on 1 April 1994. (The regulator established one cap based on a simple average of pool prices over the fiscal year and a higher cap on the average of pool prices weighted by demand levels.)

Although the pool system provides an effective means of clearing prices, it may not be effective in encouraging cost reductions. Moreover, the bulk of the generator's sales are covered by contracts with their customers, called 'contracts for differences': those contracts are essentially financial instruments providing hedges against fluctuations in the pool prices. Although the producer will receive the PIP from the pool and the buyer will pay the POP to the pool, they will compensate each other for the difference with the price stipulated in the contract. Lastly, the effectiveness of the pool system in providing the right incentives to invest in new generating plants (particularly regarding their location, an important factor given the technical constraints specific to electricity networks), remains to be established.

The primary bottleneck in the gas industry, too, is transportation (pipelines). And, as in the electricity industry, incumbent operators of the bottleneck have traditionally been vertically integrated into production and sometimes into distribution.

Historically, pipelines have served (and still do in most countries) as gas merchants. They would buy gas internally or externally at the wellhead and sell it at the city gate to local distribution companies or industrials. In economic parlance, the bottleneck segment has thus traditionally been located 'downstream', which means that prices to customers (distribution companies, industrials) are set by the pipeline, while the upstream located potentially competitive segment (production) just supplies an input in the production process.

<sup>(1)</sup> In Scotland the industry has been divided into three vertically integrated companies: Scottish Power and Scottish Hydro-Power belong to the private sector while Scottish nuclear has remained in the public sector, like Nuclear Electricity in Britain. In Northern-Ireland, which has no connection neither with Great-Britain nor with Ireland, the government decided in March 1992 to sell the incumbent monopoly (Northern Ireland Electricity, NIE)'s generating plants in three batches, NIE keeping transmission and distribution to become NIE(TD&S), privatised in June 1993.

<sup>(2)</sup> In December 1995 the RECs listed NGC on the Stock Exchange; it then became the National Grid Group (NGG).



In the USA the movement toward competition was set in motion by the Natural Gas Policy Act of 1978. A key step in the opening of competition was the so-called Order 436 in 1985, which instituted open access. In a sense this order reversed the industry structure. Local distribution companies and industrials were now allowed to buy gas directly from producers and to ship gas via pipelines. Integrated pipelines had to unbundle sales and transportation services, and transportation services became an input supplied (internally and externally) to gas merchants.

The order required pipelines to (a) convert firm sales entitlements to transportation equivalents (although by 1992, customers had not used this option), (b) not discriminate in their choice of access charges, and (c) design maximum rates to ration capacity during peak periods and to maximise throughput for firm service during off-peak periods.

The outcome was the development of a sizeable spot market and a sharp decrease in the role of pipelines as merchants. In 1992, independent producers accounted for 79 % of pipeline transportation. It should be noted, though, that competitors in 1992 had no open access to storage and specialised mainly in interruptible service, while pipelines were dominant in the firm service segment.

For future experiments it is useful to record some of the difficulties encountered in the US experience. (Many of these difficulties are of course common to other liberalised network industries.) On the incumbent side, pipelines are constrained by obligations to stand ready to provide gas on demand without notice. The management of the peak load period and who is accountable for firm supplies (e.g. to local distribution companies) is an important issue. Second, there have been some stranded assets (supply commitments that turned out to be uneconomical when the price of gas fell and hard to cover when the market was liberalised). Third, there is the standard issue of the design of access charges, and whether the access charges should be regulated in relation to (in extreme cases tied to) the final prices. In the USA unbundled transportation was regulated on the basis of a rate of return (cost of service) methodology; by fear of cross-subsidies from the regulated segment (pipelines) to the competitive segment (production), pipelines were forced to sell gas as merchant at cost-based rates (some weighted average cost of gas), which then created prices for internally supplied gas above

market rates. It was thus difficult for pipelines to compete in the competitive segment. On the entrant side, entrants have, as we mentioned, found it difficult to compete effectively on the firm service segment, due to their disadvantage in access to storage facilities and to the pipeline's interrupting their service when the pipeline needs the capacity for its own firm sales services.

In the UK, the 1982 Oil and Gas Act allowed, as the US Order 436 did, for access to the pipeline network of British Gas (BG) by competing producers for supplies to customers with demands over 25 000 therms per year. This (common carriage) legislation proved rather ineffective as no access agreement was signed between 1982 and 1990. The Gas Act of 1986 led to the privatisation of BG and to the creation of a regulatory agency, Ofgas. The Director General of Gas Supply may specify the access charges if parties cannot agree. These access charges should cover an appropriate proportion of operating cost, depreciation and a reasonable rate of return. BG kept its obligation to provide gas supplies when required.

The framework of the 1986 Act was criticised along several dimensions. First, it allowed a complete passthrough of gas purchase costs in the pricing formula and thus provided low incentives for BG to purchase gas at a low price. (This feature was removed in 1992, when the BG price was based on a gas price index rather than its actual cost of gas.)

Second, there was no accounting separation between regulated segments (in particular, transportation and tariff customers) and deregulated segments (non-tariff customers). In 1993, BG announced under pressure a separation between transportation and storage (TransCo) and its trading business (Business Gas); the separation was completed in 1994, with the view that TransCo would operate at arm's length with Business Gas and offer equal terms to the internal supplier (Business Gas) and to external suppliers. On the other hand, the government turned down the 1993 recommendation of the Monopoly and Mergers Commission that BG divest its supply business (which would have created vertical separation as in the UK electricity industry), although it forced BG to erect 'Chinese walls' between Business Gas and TransCo and banned contacts between certain employees.

Third, some also complained about the widespread price discrimination practised by BG for non-tariff customers. (BG was later banned from practising price discrimination in the non-tariff market.) Fourth, a widely criticised policy was the use of market share targets for entrants. Such targets may well promote collusion between BG and its competitors in a process in which BG raises its price to lower its share while the entrants also raise their prices in order not to force BG to raise its price to not exceed its allowed share.

Many of the problems faced in the USA were also encountered in the UK. First, the incumbent faced restraints on competition for large customers (redefined until April 1996 as customers with demand exceeding 2 500 therms). BG was obliged to publish price schedules for medium size firm customers and interruptible customers, give 21 day public notice for changes and not change prices more than once a month (these obligations were suspended in June 1995). This may have resulted in misallocation in production. Second, BG (or rather now its trading arm, Business Gas) was saddled with expensive inherited gas purchase contracts. Third, entrants were handicapped by the large number of existing contracts. In its 1991 Gas Review, the Office of Fair Trading recommended that BG release some of its contracting gas, and BG agreed to reduce its share of the non-tariff market to 40 %, and offered contracted gas for sale. Fourth, there has been a heated debate about the rate of return allowed for transportation and thus about the access charge (see Armstrong et al. (1994) and the OFGAS reports). And competitors complain about TransCo's poor service quality. More recently, BG rejected Ofgas proposals for transportation and storage price control, and Ofgas referred the matter to the Monopolies and Merger Commission, which should publish its findings in April 1997.

Two recent developments in the UK gas industry are noteworthy. The vertical separation issue raised by the Monopolies and Mergers Commission in 1993 has resurfaced, in a different guise. In February 1996, BG announced its intention to pursue a demerger and to create two separate companies (the divestiture was entering its final phase as of February 1997). BG will be broken into two firms. Centrica will be a holding for the UK domestic gas distribution business and will also include the British Gas service and retail businesses. (The future) Centrica currently serves more than 19 million households but will face tough competition once the domestic market is opened fully to competition

in 1998. The competition should come from the regional electricity companies as well as from some big oil companies. BG, the second company, will own the pipeline system (TransCo) and the international oil and gas exploration and production business.

The second recent development has been the renegotiation of the 'take or pay' contract. Under these contracts (worth tens of billions of pounds) BG must pay for gas, even though competition substantially reduces its market share and thus its demand for the input. (BG's share of the commercial market open to competition has plummeted to about a third.) BG is looking for or entering agreements with its suppliers (Shell, ESSO, Mobil, BP, etc.) to pay cash or transfer assets (the Morecambe gas fields) in exchange for renegotiated supply contracts.

#### **4. Competition between infrastructure operators and the problem of two-way access**

##### **Introduction**

We now turn to situations where several infrastructure operators are competing head-to-head, in an unregulated environment, for final customers. A first issue arising in this context concerns the compatibility of the competing networks, and has been already discussed above. We now turn to a second and related issue: even when network operators choose or are required to be compatible or interconnected, there remains the issue of the determination of the mutual access conditions provided by the competing operators. We now discuss the determination of these access conditions in an unregulated environment, and analyse their welfare effects. We will assume in this section that each network provides a full range of services.

The developing liberalisation we observe in many network industries has already produced, or is likely to produce, substantial changes in the way a number of the industries covered by this report operate. In particular, in some of those industries, regulatory scrutiny has already given way (or will soon) to a competitive market place from which detailed regulation has withdrawn. This is for example the case for air transportation, which was liberalised long ago in the USA and will soon be in Europe. Similarly, in telecommunications, legislators, regulators, and antitrust authorities envision a transitional period followed by the substitu-

tion of competition policy for regulatory supervision. This view is for example expressed in the United Kingdom by OFTEL's consultative document of December 1994 'A Framework for Effective Competition', as well as in the USA by the February 1996 Act passed by Congress, that will facilitate entry (including by the long distance companies) into the regional Bell Companies' territories in order to enhance competition in local phone service while allowing the latter to enter the long distance market once 'sufficient competition has developed in the local phone market'.<sup>(1)</sup> Similarly, in Europe the industry will be fully open to competition by January 1998. The New Zealand's 'light-handed' regulatory regime, which relies on private negotiations between competitors to secure interconnection agreements, provides in that respect a test laboratory, since the regulatory authority has been abolished and the dominant firm (Telecom) and the entrants (Clear, BellSouth New Zealand) are subject only to general antitrust provisions on the abuse of dominant position and on vertical restraints<sup>(2)</sup>.

Two main concerns arise in this context. First, network externalities can be achieved only if operators are interconnected. The provision of mutual access requires some form of cooperation among competitors, who must agree on modes of operation and especially on access prices; one may then fear that established networks use their interconnection agreements to facilitate cooperation in the final market as well (collusive behaviour concern). Second, it is often suggested that, during a transition period toward competition, often characterised by the presence of a large, well-established dominant network operator, entrants may be handicapped by the incumbent's reluctance to provide access to its network on a reciprocal basis and at a reasonable price (entry deterrence concern)<sup>(3)</sup>. We will

consider those two issues in turn, by discussing successively the role of access charges in mature and immature phases of network competition<sup>(4)</sup>.

For the sake of presentation, we will refer to telecommunications networks, where interconnection agreements are most common. It should be acknowledged that similar issues may arise in other network industries. For example, interconnections between transportation networks suppose some form of coordination on schedules and route choices, and may involve some agreement on prices as well. Payment cards and ATM networks also involve some cooperation among banks in the determination of 'access fees' (called 'interchange fees'). Yet, of the network industries covered by this report, the telecommunications industry is so far the one on the verge of, or already, experiencing infrastructure competition. The telecommunications industry is also the industry where the economic analysis is the most advanced.

#### 4.1. Network competition in a mature industry

Consider a situation where two telecommunication operators provide a full range of services, and assume that the two operators: (i) have symmetric cost structures; (ii) have full coverage, so that all customers can choose between the two operators (this corresponds to the 'mature' industry assumption); (iii) are interconnected, so that subscribing to one operator allows to call subscribers to the other network. Typically, those two networks will compete for subscribers but, at the same time, will have to agree on mutual access conditions for cross-network communications. These access conditions will in general involve quality standards, access prices, and so on. For the sake of exposition, we will assume here 'perfect' access quality, and focus on the impact of access charges.

In practice, access charges usually consist of an interconnection price paid for each unit of communication terminating on the other network; these access charges can be set either by mutual agreement, or unilaterally (each network setting the price for access to its own network), or else can be subject to regulatory scrutiny. In the latter case, the regulatory intervention can take the form of a cost based rule (fully-distributed costs), a

(1) Cf. *Congressional Quarterly* (1995, p. 6): 'The House and Senate have passed competing bills to promote competition and deregulation in telecommunications, both of which drew broad, bipartisan support. The two proposals (HR 1555 — H Rept 104-204 Part I; S 652 — S Rept 104-23) have the same goal: to allow all telecommunications companies to compete head to head in one another's markets, with as little government regulation as possible.' An agreement between President Clinton, who threatened to veto the reconciliation Bill, was reached in December 1995. See Schwartz (1995) and the 1996 *Economic Report of the President* for good accounts of the debate on telecommunications reform in the USA.

(2) Section 36 of the Commerce Act may be invoked to limit access charges that a vertically integrated dominant operator can charge for the essential input.

(3) In New Zealand, bitter disputes between the incumbent and the entrants suggest that satisfactory, freely negotiated interconnection agreements may not come about easily. A wide ranging consultation has been launched with the aim of promoting competition with minimal State intervention. (See the joint document of the Ministry of Commerce and the Treasury (1995).)

(4) See Laffont, Rey and Tirole (1996a,b) for a detailed analysis. Armstrong (1996) independently provided some of the insights.



price cap, or of regulatory rules such as reciprocity, ECPR, etc.

We will first consider the case of a given, reciprocal access charge and examine (retail) price competition given this common access charge. We will discuss afterwards the impact of alternative ways of setting access charges. Also, retail price competition can take different forms: uniform linear prices, non-linear prices, or discriminating prices for which the price of a call depends on whether the call terminates on the same or the rival network. We successively discuss these various alternatives.

#### *4.1.1. Uniform price competition*

Let us thus assume that access charges are given and symmetric, and furthermore assume here that networks compete in linear prices. In such a context, when choosing the retail price offered to its subscribers, each operator will take into account the impact of those prices on: (i) its market share, that is, on the number of customers willing to subscribe to its own network (market share effect); (ii) the revenue generated by those subscribers (retail revenue effect); and (iii) the access revenue (or deficit) generated by cross-networks communications (access revenue effect). An increase in the final price generally decreases the market share (market share effect), but increases the retail revenue generated by each customer (retail revenue effect), as long as the retail price does not exceed the monopoly level; however, since such a price increase decreases the volume of calls generated by the network's subscribers (while having no impact on the calls received from the rival's subscribers), it is likely to also increase the access revenue (or decrease the access deficit) generated by cross-networks communications (access revenue effect). Whereas the first two effects illustrate the classic trade-off common to any form of oligopolistic competition, the latter effect is specifically due to the presence of access charges, and gives further incentives to increase retail prices. This access revenue effect is higher when access charges are high, but also when market shares are close to each other (so that the proportion of cross-networks communications is high).

This access revenue effect creates a factor of instability of competition between the networks. Indeed, if the access charge is high, this effect pushes the final price upwards (possibly higher the monopoly level) in any shared market equilibrium. But when prices are very high, and if networks are sufficiently close substitutes,

each operator has an incentive to undercut its rival and corner the market. But a cornered market configuration cannot be an equilibrium either, since then: (i) either one network makes a positive profit, in which case the other could mimic it and obtain half of this profit; or (ii) no network makes a profit, in which case a network could obtain positive profit by raising its price and generating access revenue (by charging a price high enough, it gets a small market share but induces its own subscribers to call less, and thus generates an access surplus).

But this access revenue effect makes also the access charge an instrument of tacit collusion. As just noted, because of this effect, at a shared-market equilibrium the retail equilibrium price increases with the access charge. Indeed there is a level for the access charge (decreasing with the fixed cost of connection and increasing with the substitutability of the networks) which induces the monopoly price as a non-cooperative or 'competitive' equilibrium (see Laffont, Rey and Tirole (1996a)). In contrast, the access price which would be necessary to induce socially optimal prices (namely, Ramsey prices) is smaller than the marginal cost of access.

#### *4.1.2. Competition in two-part tariffs*

Assume now that competing operators can offer two-part tariffs to their subscribers. The existence of a fixed fee enables them to charge a lower marginal price. Indeed, they find it optimal to set a marginal price that reflects their own perceived marginal cost, using the fixed (subscriber) fees to recover any share of consumer surplus that is compatible with competitive pressure. High access charges still lead to high marginal retail prices: the access charge, which is incurred for all outgoing, or off-net, communications, still affects on average the marginal cost of the internal subscribers' communications; hence, high access charges imply high marginal costs (on average) and thus again high marginal retail prices. The tension between excessive marginal prices and the temptation of undercutting thus remains and still generates, for high access charges and high substitutability, an instability of the competition between the two networks.

However, the impact of the access charge on the intensity of competition is a priori less obvious than in the case of uniform price competition. Since marginal prices now simply reflect marginal costs, it only remains to analyse the determination of fixed fees. But whereas an increase in the fixed fee adversely affects the market share and raises the retail revenue from the

subscribers, the impact on the access revenue is a priori less clear. In particular, in contrast with the marginal price charged to the subscribers, the fixed fee does not affect the volume of calls. It may still affect the access revenue through changes in relative market shares, which affect the volume of cross-networks communications. Such an effect however does not arise if for example the networks generate as many calls per customer and each customer is as likely to call subscribers from both networks. In that case, the access revenue effect disappears and, since this effect is the only one directly related to the access charge, the access charge no longer is a collusive device. An increase in the access charge then mainly leads to an increase in the marginal retail price, reflecting the increase in the marginal cost perceived by each operator, and thus to a decrease in consumption, but does not affect the operators' profit <sup>(1)</sup>. (Of course, all the traditional reasons which make non-linear prices difficult to design efficiently (incomplete information on consumers' tastes, negative redistributive effects of fixed fees), to the extent that they cannot be overcome by menus of tariffs, are likely to restore partially the tacit collusion effect of high access charges.) Note lastly that, although marginal retail prices are always lower with two-part tariffs than with linear prices, since two-part tariffs eliminate or at least reduce the scope for double marginalisation, equilibrium profits need not be lower than under uniform pricing: they are likely to be lower if the access charge is close to the level which sustains monopoly prices with uniform pricing competition, but may be higher otherwise.

#### 4.1.3. Price discrimination based on call termination

If networks are allowed to charge different on-net and off-net prices, they will do so since their perceived marginal costs are different for those two types of communications: an operator incurs the true marginal cost for on-net calls, but has to pay an access mark-up or subsidy on off-net calls. As a result, the price for off-net calls is higher than for on-net calls, assuming for the purpose of the analysis that the access charge exceeds the marginal cost of access. Allowing for such termination-based discrimination does not however suffice to

eliminate the already noted instability of price competition: high access charges still lead to high (off-net) prices in any shared-market equilibrium, which again creates an incentive to undercut the rival and corner the market; and cornered-market equilibria, as earlier, cannot be sustained. So again there may be instability for high access charges.

Whereas higher access charges clearly lead to higher off-net prices, the impact on on-net prices is less obvious. First, note that the access revenue effect described above now concerns off-net prices, and not on-net prices. Thus the access charge has no direct effect on the determination of on-net prices. However, since an increase in the access charge leads to higher off-net prices, it indirectly alters the first two effects, in a way that may well lead to a decrease in on-net prices; the wedge between on-net and off-net prices is detrimental to consumption efficiency, but may intensify competition, with ambiguous welfare effects <sup>(2)</sup>. In particular, if the two networks are poor substitutes and if there is a mark-up on access, social welfare may be higher under price discrimination than under uniform pricing.

Building on this analysis, we can now briefly discuss the case where the operators are allowed both to offer two-part tariffs and to price discriminate according to where calls terminate. Marginal prices then reflect again marginal costs: thus the marginal price for on-net calls equals the true marginal cost, whereas the marginal price for off-net calls equals the perceived marginal cost, including the access mark-up. The presence of a mark-up in the access charge thus again imposes an inefficiency, since it creates a wedge between on-net and off-net marginal prices (note that this wedge is based on the difference between perceived marginal costs, not actual marginal costs). The analysis is otherwise similar to the case of two-part tariffs without price-discrimination: when setting the fixed fee, an operator takes into account the impact on its market-share and on the revenue generated from its customers but, since marginal prices include incurred access charges, it does not consider any impact on the access revenue (except, possibly, for the impact on relative market shares, which may affect how many calls it receives from the other network). Hence again, the access charge cannot be used as an instrument device <sup>(3)</sup>.

<sup>(1)</sup> For example, Laffont, Rey and Tirole (1996a) shows that, for horizontally-differentiated networks with isotropic calling pattern and iso-elastic demands, equilibrium profits are exactly independent of the access charge: an increase in the access charge only results in a decrease in consumption and consumer surplus, due to higher retail marginal prices, while fixed fees are adjusted so as to maintain constant the operators' profits.

<sup>(2)</sup> See Laffont, Rey and Tirole (1996b).

<sup>(3)</sup> A local increase of the access charge may even reduce equilibrium profits; see Laffont, Rey and Tirole (1996b).

#### 4.1.4. The determination of access charges

We have assumed so far that access charges were reciprocal and exogenously given. We now analyse various ways to set those access charges. From a normative point of view, the best choice would consist in an access charge leading to retail prices equal to the Ramsey price, i.e. the lowest price that is compatible with the operators' budget constraints. (Note that, in themselves, access charges only serve a redistributive purpose and thus have no direct impact on total economic welfare <sup>(1)</sup>.) The socially optimal level of the access charge depends on the nature of retail competition: as already mentioned, under uniform pricing competition, the reciprocal access charge leading to the Ramsey prices is lower than the access marginal cost. In contrast, if the operators can discriminate among customers via, say, perfectly tailored two-part tariffs, then, since marginal retail prices reflect marginal costs, it is socially optimal to set access charges equal to the access marginal cost.

These desirable levels for the access charges are likely to differ from those that the operators would like to choose. If they could agree on mutual access charges, then under uniform pricing competition (or provided that two-part tariffs and other non-linear retail pricing policies do not allow the operators to perfectly discriminate among their subscribers), the operators would incorporate a mark-up in their access charges, since higher access charges weaken retail competition and lead to higher retail prices (as mentioned, a high enough access charge even leads the 'competing operators' to set retail prices at the monopoly level). Note that if access charges are determined non cooperatively, the situation can get even worse, because of a standard double marginalisation problem: each operator has an incentive to increase its access charge over the level that would sustain monopolistic retail prices, since it fails to take into account the resulting reduction in its rival's profits. This double marginalisation problem is clearest when the networks do not compete for market share, that is, when networks are very differentiated. But as substitution increases, and competition for customers exerts a downward pressure on the retail price,

this competition may also induce networks to jack their access price up in order to force their rival to raise its retail price and lose market share.

Last, let us briefly examine the implications of alternative regulatory interventions such as the efficient component pricing rule (ECPR). This rule imposes a ceiling on the access price charged by a network to its competitors, that depends on the final price charged by the network on the competitive segment: the access price charged by the network should not exceed the network's price on the competitive segment minus the network's cost on that segment. In other words, the access charge should not exceed the opportunity cost for the network of losing a call on the retail segment. An alternative approach to ECPR is the 'imputation methodology', according to which a bottleneck owner should be required to provide monopoly service elements at the price it imputes to its own competing services (subject to a budget constraint on the competitive segment, to be interpreted as resulting from 'accounting separation'), that is the difference between retail price and marginal cost on the retail segment <sup>(2)</sup>.

It should first be pointed out that ECPR has no obvious interpretation in 'two-way bottlenecks' situations such as discussed here. In particular, while there is no difference between the opportunity cost and the imputation methodologies in the standard context in which they are applied (single bottleneck giving access to a potentially competitive segment), their philosophies and implications differ when applied to network competition. Also, the very notion of 'bottleneck' may vary according to whether an *ex post* or an *ex ante* perspective is adopted. *Ex post*, that is after consumers have joined a network, both the originating and terminating ends are bottlenecks; each is monopolised by a network and is essential to the completion of a call. *Ex ante*, that is before

<sup>(1)</sup> In particular, in any 'balanced' competitive equilibrium, each operator receives as much as he pays its rival, and the access revenue is zero as long as the access charges are reciprocal. Nevertheless, as stressed by the above analysis, the level of this reciprocal access charge still affects the competitive behaviour of the operators, and thus has a real impact on economic welfare.

<sup>(2)</sup> See Hausman (1994a,b). The *Clear-Telecom* decision of New Zealand's highest court (the Privy Council of the House of Lords in London) provides a good example of the application of the ECPR. In its dispute with Clear Communications, New Zealand Telecom, the dominant operator, argued that it was entitled to recover its opportunity cost when setting its interconnection charge. While the Court of first hearing had ruled in favour of Telecom and the Court of Appeal in Clear's favour, the Privy Council finally decided that the use of ECPR by Telecom was lawful under section 36 of the Commerce Act. This endorsement of ECPR by the Privy Council has been controversial, if only because the rule was designed to apply to a regulated environment (as William Baumol pointed out in testimonies), whereas little is known about the impact of ECPR in an industry where retail prices are not regulated. (See Laffont and Tirole (1994, 1996) for an assessment of the optimality of ECPR under retail price regulation, and Baumol et al. (1996) for a further discussion.)

consumers have joined a network, the complementary segments are (imperfectly) competitive. It is thus not clear how ECPR is likely to be interpreted by regulators and courts in the context of network competition, and then investigate its impact on competition <sup>(1)</sup>.

A first inspection seems however to suggest that ECPR may soften price competition. Suppose first that the networks agree on an access price and then pick their retail prices subject to ECPR: then the networks can sustain monopoly prices by initially agreeing to a high access charge, namely, by choosing a mark-up in the access charge that exactly reflects the monopoly retail mark-up; in effect, ECPR then actually prevents the operators from lowering their prices below the monopoly level. In other words, in the case of an *ex ante* agreement, a high access charge is under ECPR a commitment to charge a high retail price. Suppose now that the two networks set their access and retail prices simultaneously. Then under ECPR, setting a high access charge requires setting a high retail price and therefore facing the prospect of a low market share. A network thus cannot undercut and gain market share without providing its competitor with a windfall gain on access.

#### 4.2. Entry in network competition

We now discuss the issue of entry in a transition period. As mentioned, the main concern here is that an already well-established network operator may be reluctant to provide potential entrants with access to its network at a reasonable price.

Note that entry can take several forms. For example, the entrant may either lease facilities from the incumbent or build its own facilities. The second option involves a fixed joint and common cost of partially or fully duplicating the incumbent's network <sup>(2)</sup>. We analyse in turn resale-based and facilities-based entry.

Let us first assume that the entrant leases its facilities from the incumbent. In that case, the incumbent's pricing policy must take into account the incumbent's joint and common cost. This common cost can a priori be recovered through an increase in the access charge, or through the price at what the entrant leases the incumbent's facilities (the lease price). As stressed above, in the absence of joint and common cost some subsidy on access charges must be built in order to offset the networks' market power and to keep the unregulated retail prices down. This downward pressure on retail prices can now be obtained in two alternative ways: a low interconnection charge or a low lease price, which would subsidise the entrant and thus encourage both the incumbent and the entrant to build market share. As shown in Laffont, Rey and Tirole (1996a), it is however better to create the downward pressure on retail price through the interconnection charge, as this preserves the level playing field between the two operators; in contrast, a subsidy on the lease price would favour the entrant to the detriment of the incumbent and would thus generate excessive entry.

We now turn to facilities-based entry. To fix ideas, assume that there are two network operators, a well-established one, with full coverage, and another, the entrant, which has initially no coverage and must incur an investment cost related to the coverage it is planning to offer.

It is well-known that a 'large' network has little incentive to make itself compatible with other smaller ones; similarly here, if the entrant's coverage is small, the incumbent network has an incentive to refuse interconnection, since in the absence of interconnection it can corner the market at a profitable price, maybe even at the monopoly price. Hence, if interconnection agreements are freely negotiated, the incumbent will then have an incentive to delay indefinitely the reach of an agreement. (On the other hand, if an agreement has to be reached, the entrant has an incentive to over-invest in coverage in order to reduce the incumbent pre-agreement profit and thus to reach a better deal.)

Similarly, if interconnection is mandated but each operator is left free to set its access charge, the incumbent has an incentive to set its own access charge at a prohibitive level, as this constitutes a standard 'raising rival's cost' strategy.

<sup>(1)</sup> See Laffont, Rey and Tirole (1996a) for a detailed analysis of possible adaptations of the ECPR to the case of competing networks.

<sup>(2)</sup> The February 1996 Telecommunications Act and the August 1996 FCC Report and Order envision three types of entry in the local market. Under resale entry the incumbent leases the entire local loop to entrants. Under facilities-based entry the entrants build their own local network. Last, under unbundling (an hybrid arrangement), the entrants lease some to the incumbent's facilities (say, their transmission facilities) and build some others themselves (say, the switches). Under facilities-based entry the only access charge is the transport-and-termination or interconnection charge considered previously; in contrast, resale or unbundling entry involves an access charge paid by the entrants to the incumbent for elements or the whole of the local network, on top of the transport and termination charges that the incumbent and the entrants pay to each other.

This suggests that imposing reciprocity in the setting of access charges may be a particularly good idea when entry or coverage is at stake, that is, when the industry is still in an immature phase of network competition. However, reciprocity will not in general suffice to prevent entry deterrence. If for example the incumbent can discriminate between its monopoly territory and the competitive territory, it can still effectively block entry even under a reciprocity rule by insisting on a high access charge, and a high retail price for off-net communications. In effect, by discouraging its own subscribers to call the other network (a strategy the entrant cannot follow if its coverage is small), it avoids paying high access charges to its rival. Hence, even though formally the access charges are reciprocal, termination-based price discrimination allows the incumbent to block entry as effectively as if it could freely set its own access charge: a high access charge has basically the same impact as a lack of interconnection.

A reciprocity rule for the access charge, coupled with a ban on termination-based discrimination, may still not suffice to generate the most desirable outcome: if the access charge is reciprocal, and the incumbent cannot price discriminate between its monopoly territory and the competitive territory, it is easy to check that the entrant has an incentive to under-invest in coverage (or other dimensions of quality) in order to soften competition. The reason is that when the entrant has a small coverage, the incumbent has a large captive market, which induces him to maintain high prices. Hence, if a reciprocal access price is mandated, the entrant may not be handicapped by its smaller coverage but it may indeed elect to maintain a small coverage for strategic reasons.

We conclude that the setting of a reciprocal access charge by an incumbent firm can be a powerful instrument to blockade entry as long as entrants cannot quickly achieve a high coverage of the market.



# References

- Aghion, P. and Tirole J. (1997), 'Formal and Real Authority in Organisations', *Journal of Political Economy*.
- Alchian, A. A. (1965), 'Some Economics of Property Rights', *Il Politico*, 30, pp. 816–829.
- Ali, A. I., Lerme, C. S. and Seiford, L. M. (1995), Components of Efficiency Evaluation in Data Envelopment Analysis, *European Journal of Operation Research*, 80, pp. 462–473.
- Areeda, P.E. (1985), *Antitrust Law*, Boston: Little, Brown and Company.
- Arid, P. (1981), *Antitrust Analysis*, Boston: Little, Brown and Company, third edition.
- Armstrong, M. and Doyle, C. (1994), 'Access Pricing, Entry and the Baumol–Willig Rule', DP 9422, University of Southampton.
- Armstrong, M. and Vickers, J. (1991), 'Welfare Effects of Price Discrimination by a Regulated Monopolist', *Rand Journal of Economics*, 22, pp. 571–580.
- Armstrong, M. and Vickers, J. (1995), 'The Access Pricing Problem', mimeo, University of Southampton and Oxford University.
- Armstrong, M. (1996), 'Network Interconnection', mimeo.
- Barla, P. and Perelman, S. (1989), 'Technical Efficiency in Airlines under Regulated and Deregulated Environment', *Annals of Public and Cooperative Economics*, 60, pp. 103–124.
- Bauer, P.W. (1990), 'Recent Developments in the Econometrics Estimation of Frontiers', *Journal of Econometrics*, 46, pp. 39–56.
- Baumol, W. J. (1993), 'Deregulation and Residual Regulation of Local Telephone Service', AEI Studies in Telecommunications Deregulation, American Enterprise Institute for Public Policy Research, New York University.
- Baumol, W. J., Panzar, J. and Willig R. (1982), *Contestable Markets and the Theory of Industry Structure*, NY: Harcourt Brace, Jovanovich.
- Baumol, W. J., and Sidak, G. (1994a), *Toward Competition in Local Telephony*, Cambridge: MIT Press.
- Baumol, W. J., and Sidak, G. (1994b), 'The Pricing of Inputs Sold to Competitors' *Yale Journal on Regulation*, pp. 11, 171–202.
- Baumol, W., Ordover, J., and Willig, R. (1997), 'Parity Pricing and its Critics,' *Yale Journal on Regulation*, 14, pp. 145–163.
- Bernstein, M. (1955), *Regulating Business by Independent Commission*, Princeton, NJ: Princeton University Press, 105, pp. 1–29.
- Boiteux, M. (1956), 'Sur la Gestion des Monopoles Publics Astreints à l'Equilibre Budgétaire', *Econometrica*, 24, pp. 22–40.
- Borenstein, S., MacKie-Mason, J. K. and Netz, J. S. (1995), 'Antitrust Policy in Aftermarkets', *Antitrust Law Journal*, 63, pp. 455–482.
- Bork, R. (1978), *Antitrust Paradox*, New York: Basic Books.
- Burgat, P. and Jeanrenaud, C. (1990), 'Mesure de l'Efficacité Productive et de l'Efficacité-Coût: Cas des Déchets Ménagers en Suisse', Working paper No 9002 de l'Institut de Recherches Economiques et Régionales, Université de Neuchâtel.

- California PUC (1994), 'In the Matter of Alternative Regulatory Frameworks for Local Exchange Carriers', Decision 94-09-065, Public Utilities Commission of the State of California.
- Cave, M. (1993), 'Interconnection Issues in UK Telecommunications', paper presented at the conference on 'Access pricing in regulated industries', London, November 12, 1993.
- Cazals, C., de Rycke, M. Florens, J.P. and Rouzaud, S. (1997), 'Scale Economies and Natural Monopoly in the Postal Delivery: Comparison Between Parametric and Non Parametric Specifications', in M. A. Crew and P.R. Kleindorfer, editors, *Managing Change in the Postal and Delivery Industries*, Boston: Kluwer Academic Publishers.
- Chemla, G. (1995), 'Competition, Investment, and Vertical Integration', mimeo, Financial Markets Group, London School of Economics.
- Congressional Quarterly (1995), 'Special Report. The Fine Print: A Side-by-Side Comparison of the House and Senate Telecommunications Bills', Washington, D.C., September 23.
- Cote, D. (1989), 'Firm Efficiency and Ownership Structure. The Case of US Electric Utilities Using Panel Data', *Annals of Public and Cooperative Economics*, 60, pp. 431–450.
- Cowie, J., and Riddington, G. (1996), 'Measuring the Efficiency of European Railways', *Applied Economics*, 28, pp. 1027–1035.
- Crémer, J. (1995), 'Arm's-length Relationships', *Quarterly Journal of Economics*, 2, 275–295.
- Crew, M. and Kleindorfer, P. (1994), 'Pricing, Entry, Service Quality and Innovation under a Commercialized Postal Service', Chapter 6 in *Governing the Postal Service*, G. Sidak, ed., AEI Press, Washington.
- Cubbin, J., Domberger, S. and Meadowcroft, S. (1987) 'Competitive Tendering and Refuse Collection: Identifying the Sources of Efficiency Gains', *Fiscal Studies*, 8, pp. 69–87.
- Curien, N., Jullien, B. and Rey, P. (1993), 'Pricing Regulation under Bypass Competition', mimeo, CNAM and LEI.
- Delhausse, B., Fecher, F. Perelman, S. and Pestieau, P. (1991), 'Measuring Productive Performance in the Non-life Insurance Industry: The Case of French and Belgian Markets', mimeo.
- Dewatripont, M., and Tirole, J. (1995), 'Advocates', mimeo, ULB, Brussels, and Institut d'Économie Industrielle, Toulouse.
- Dewatripont, M., and Tirole, J. (1996), 'Advocates', Nota di lavoro, ETM 96/43, Fondazione Eni Enrico Mattei, Milan.
- Distexhe, V. (1993), 'L'Efficacité Productive des Services d'Enlèvement des Immondices en Wallonie', mimeo, Cahiers Économiques de Bruxelles, 137, pp. 119–138.
- Engel, E., Fischer, R. and Galetovic, A. (1997), 'Highway franchising : pitfalls and opportunities', *American Economic Review*, 87(2), May.
- Ergas, H. and Ralph, E. (1994), 'The Baumol-Willig Rule: The Answer to the Pricing of Interconnection?', mimeo, Trade Practices Commission, Canberra.
- Färe, R., Grosskopf, S. Logan, J. and Lovell, C.A.K. (1985), 'Measuring Efficiency in Electric Power Generation due to Environmental Controls', *Resource and Energy*, 8, pp. 167–184.
- Färe, R., Grosskopf, S. and Lovell, C.A.K. (1994), *Production Frontiers*, New York: Cambridge University Press.
- Farrell, M. J. (1957), 'The Measurement of productive Efficiency', *Journal of the Royal Statistical Society, Series A*, 120, pp. 253–281.
- Fecher, F., Kessler, D. Perelman, S. and Pestieau, P. (1993), 'Productive Performance of the French Insurance Industry', *Journal of Productivity Analysis*, 4, pp. 77–93.
- Filippini, M. and Maggi, R. (1991), 'Efficiency and Regulation in the Case of the Swiss Private Railways', *Journal of Regulatory Economics* (forthcoming).



- Fiorina, M. (1985), 'Group Concentration and the Delegation of Legislative Authority', in *Regulatory Policy and the Social Sciences*, ed. Roger G. Noll, Berkeley: University of California Press.
- Florens, J. P., Ivaldi, M. and Larribeau, S. (1996), 'Sobolev Estimation of Approximate Regressions', *Econometric Theory*, 12, pp. 753–772.
- Gathon, H. J. and Pestieau, P. (1992), 'Faut-il encore mesurer la performance des entreprises publiques?', *Annals of Public and Cooperative Economics*, 63, pp. 621–644.
- Good, H. D., Roeller, L. H. and Sickles, R. (1991a), 'EC Integration and the Structure of The Franco-American Airline Industries: Implications for Efficiency and Welfare', in W. Eichorn et al. (eds), *Models and Measurement of Welfare and Inequity*, Heidelberg (forthcoming).
- Good, H. D., Roeller, L. H. and R. Sickles (1991b), 'Airline Efficiency Differences between Europe, and the US Implications for the Pace of EC Integration and Domestic Regulation', mimeo.
- Grosskopf, S. and Vladamis, V. (1987), 'Measuring Hospital Performance: a Non-Parametric Approach', *Journal of Health Economics*, 6, pp. 89–107.
- Grout, P. (1996), 'Structure of the Price Cap: Arguments for Global Price Cap and Answers to OFTEL's Questions', mimeo, University of Southampton.
- Hancher, L. (1995), 'European Competition Law Implications for Access Pricing. Report for European Commission DGIV', mimeo, Erasmus University, Rotterdam.
- Hart, O. and Tirole, J. (1990), 'Vertical Integration and Market Foreclosure', *Brookings Papers on Economic Activity, Microeconomics*, 205–285.
- Hausman, J. (1994a), 'Proliferation of Networks in Telecommunications: Technological and Economic Considerations', mimeo, MIT.
- Hausman, J. (1994b), 'Testimonies before the Public Service Commission of Maryland'.
- Hjalmarsson, L. and Veiderpass, A. (1991), 'Productivity in Swedish Electrical Retail Distribution', University of Gothenburg, mimeo.
- Hogan, W. (1994), 'Efficient Pricing and Transmission Rights in Support of a Competitive Electricity Market', FERC Notice of Technical Conference and Request for Comments, Docket No RM93-19-00U.
- Ivaldi, M., Monier-Dilhan, S. and Simioni, M. (1995), 'Stochastic production Frontiers and Panel Data: A Latent Variable Framework', *European Journal of Operation Research*, 80, pp. 534–547.
- Kahn, A., and Taylor, W. (1994), 'Comment on Baumol and Sidak's Essay' *Yale Journal on Regulation*, 11, pp. 225–240.
- Kay, J. A. and Thompson, D. J. (1986), 'Privatisation: A Policy in Search of a Rationale', *Economic Journal*, 96, pp. 18–32.
- Laffont J. J. and Meleu, M. (1996), 'Mutual Supervision and Collusion', mimeo, IDEI.
- Laffont J. J. and Tirole, J. (1986), 'Using Cost Observation to Regulate Firms', *Journal of Political Economy*, 94, pp. 614–641.
- Laffont, J.J. and Tirole, J. (1990), 'Bypass and Cream-skimming', *American Economic Review*, 80, pp. 1042–1061.
- Laffont J. J. and Tirole, J. (1993), *A Theory of Incentives in Procurement and Regulation*, Cambridge and London: MIT Press.
- Laffont, J. J., and Tirole, J. (1994) 'Access Pricing and Competition', *European Economic Review*, 38, pp. 1673–1710.
- Laffont, J. J., and Tirole, J. (1996a), 'Global Price Caps and the Regulation of Interconnection', mimeo, IDEI, Toulouse.
- Laffont, J. J., and Tirole, J. (1996b), 'Creating Competition through Interconnection: Theory and Practice', *Journal of Regulatory Economics*, 10: pp. 227–256.

- Laffont, J. J., Rey, P. and Tirole, J. (1996a) 'Network Competition: I. Overview and Non-discriminatory Pricing', forthcoming, *Rand Journal of Economics*.
- Laffont, J. J., Rey, P. and Tirole, J. (1996b), 'Network Competition: II. Price Discrimination', forthcoming, *Rand Journal of Economics*.
- Leibenstein, H. (1966), 'Allocative Efficiency vs. X-Efficiency', *American Economic Review*, 56, pp. 392–415.
- Lenoir, N. (1991), *Optimal Structure of Regulatory Agencies facing the Threat of Political Influence*, Master's Thesis in the Science of Transportation, Massachusetts Institute of Technology.
- Manzini, A. (1990), Efficacité Publique et Privée: Analyse Théorique et Vérification Empirique dans le Cas du Transport Aérien, Thèse de Doctorat, Université de Genève.
- Masmoudi, H., and Prothais, F. (1994), 'Access Charges: An Example of Application of the Fully Efficient Rule to Mobile Access to the Fixed Network', mimeo, Ecole Polytechnique and France Télécom.
- McAfee, R. P. and Schwartz, M. (1993), 'Opportunism in Multilateral Vertical Contracting: Nondiscrimination, Exclusivity, and Uniformity', *American Economic Review* 84(1), pp. 210–230.
- Mitchell, B. and Vogelsang, I. (1991), *Telecommunications Pricing: Theory and Practice*, Cambridge: Cambridge University Press.
- Mobille, J. (1992), 'Access Charges in the United States and their Effect on the Telecommunications Market', mimeo, France Télécom, Direction du Plan et de la Stratégie.
- Neven, D., Nuttall, R. and Seabright, P. (1993), *Merger in Daylight*, London, CEPR.
- New Zealand Ministry of Commerce and Treasury (1995), 'Regulation of Access to Vertically-Integrated Natural Monopolies', Discussion Paper.
- Noll, R. G. (1989), 'Telecommunications Regulation in the 1990s' in *New Directions in Telecommunications Regulation Policy*, Vol. 1, P. Newberg, ed. Duke University Press, Durham and London.
- O'Brien, D. P. and Shaffer, G. (1992), 'Vertical Control with Bilateral Contracts', *Rand Journal of Economics* 23(3), pp. 299–308.
- OFTEL (1994), 'A Framework for Effective Competition', Consultative document. December.
- OFTEL (1995), 'Pricing of Telecommunications Services from 1997 (Consultative Document Issued by the Director General of Telecommunications)', London.
- Oum, T. H. and Yu, C. (1991), 'Economic Efficiency of Passenger Railway Systems and Implications on Public Policy', University of British Columbia, mimeo.
- Pestieau, P. and Tulkens, H. (1993), 'Assessing and Explaining the Performance of Public Enterprises', *Finanz Archiv*, 50, pp. 293–323.
- Posner, R. (1976), *Antitrust Law*, Chicago: University of Chicago Press.
- Public Utilities Commission of the State of California (1994), Decision 94-09-065 In the Matter of Alternative Regulatory Frameworks for Local Exchange Carriers. (This decision is partly based on I-87-11-033, filed in November 1987.)
- Ralph, E. (1994), 'A New Mechanism for Establishing Interconnect Fees', mimeo, Duke University.
- Ramsey, F. (1927), 'A Contribution to the Theory of Taxation', *Economic Journal*, 47.
- Rey, P. and Tirole, J. (1996), 'A Primer on Foreclosure', mimeo, IDEI, Toulouse.
- Rhodes, E. L. and Southwick, L. (1988), 'Comparison of University Performance Differences over Time', mimeo.
- Schmidt, P. (1986), 'Frontier Production Functions', *Econometric Review*, 4, pp. 289–328.
- Schwartz, M. (1995), 'Telecommunications Reform in the United States: Promises and Pitfalls', forthcoming in *Telecommunications and Energy in Systemic Transformation*, P. Welfens and G. Yarrow, ed., Heidelberg: Springer Verlag.

- Shapiro, C. (1995), 'Aftermarkets and Consumer Welfare: Making Sense of Kodak', *Antitrust Law Journal*, 63, pp. 483–511.
- Slade, M. (1995), 'Beer and the Tie: Did Divestiture Cause Higher Prices?', mimeo.
- Snyder, C. (1994), 'Vertical Foreclosure in the British Beer Industry', in *Buyers, Suppliers, Competitors: The Interaction between a Firm's Horizontal and Vertical Relationship*, PhD Thesis, Department of economics, MIT.
- Stigler, G. (1971), 'The theory of economic regulation', *Bell Journal of Economics and Management Science*, 2, spring, pp. 3–21.
- Tulkens, H. (1993), 'On FDH Efficiency Analysis: Some Methodological Issues and Application to Retail Banking, Courts and Urban Transit', *Journal of Productivity Analysis*, 4, pp. 183–210.
- Tye, W. (1994), 'Response to Baumol and Sidak's Essay', *Yale Journal on Regulation*, 11, pp. 203–224.
- Vickers, J. S. and Yarrow, G. K. (1991), 'Economic Perspectives on Privatisation', *Journal of Economic Perspectives*, 5, pp. 111–132.
- Vickers, J. (1995), 'Competition and Regulation in Vertically Related Markets', *Review of Economic Studies*, 62, pp. 1–18.
- Vickrey, W. (1995), 'A proposal for revising New York's subway fare structure', *Journal of the Operations Research Society of America*, 3, February, pp. 38–68.
- WIK-EAC (1994), *Network Interconnection in the Domain of ONP*. Final Report for DG XIII of the European Commission. November.
- Willig, R. D. (1979), 'The Theory of Network Access Pricing', in H. M. Trebing (ed.), *Issues in Public Utility Regulation*, Michigan State University Public Utilities Papers.
- Wilson, J. Q. (1989), *Bureaucracy: What Government Agencies Do and Why They Do It*, Basic Books.
- Yates, D. (1982), *Bureaucratic Democracy*, Cambridge: Harvard University Press.

# PART TWO

## Universal service

### I. Universal service — definition, cost and financing

#### 1. Introduction: Public service versus universal service

In this chapter we study the universal service obligation (USO), which occupies a prominent place in the policy debate regarding all major network industries in the EU. Before proceeding, it is important to clarify the scope of our investigation, and to point out its limitations. In particular, it has to be emphasised that our study focuses on the specific notion of ‘universal service’; it is not meant to provide a systematic discussion of the much more general concept of ‘public service’ which has been traditionally a major building block of public policy in many European countries.

Though related, public service and universal service are by no means synonyms. The two concepts have different contents, reflect different attitudes of the policy-makers, and lead to very different policy implications. To stress this point, and to achieve a clear-cut delimitation of the problems under investigation in this chapter, we shall first of all provide a general discussion of the notions of public and universal service and attempt to draw the dividing line between the two concepts.

##### 1.1. Public service

We have shown above (see Part one) that the major justification of public intervention in the economy are market failures, on the one hand, and redistributive considerations on the other hand.

Market failures arise, for instance, because of phenomena like public goods, externalities, non-convexities in production sets (increasing returns to scale), non-competitive behaviour, the incompleteness of markets or informational imperfections (especially in insurance and financial markets). They imply, that markets alone may not be sufficient to provide an efficient provision of all the goods in the economy.

Redistributional considerations provide another major justification for public intervention. They can explain, for instance, the design of the tax system (progressivity) or provide a rationale for some transfers (income maintenance programs). However, in a second-best world, they can also justify more drastic and direct public interventions which go beyond the traditional instruments of redistributive policies (see Part one, as well as the discussion below for more details).

The recognition of the need to correct or supplement the market mechanism for reasons of efficiency or equity has lead, in many countries, to the emergence of the notion of ‘public service’, according to which some activities should be directly entrusted to public authorities (or at the very least be subject to a tight regulation). This tradition is particularly well-established in countries like France and Belgium, where the concept has received a precise definition through the administrative law. Many other European countries also have a long tradition of public service even though their law does not recognise an exact counterpart to the French notion of *service public*.

The scope of this notion of public service is rather large and, depending on the specific country, it may include a more or less significant number of goods, services and activities. First of all, there is the provision by public authorities (or administrations) of ‘essential’ public goods like national defence, police protection, justice, etc. While these activities are not necessarily referred to as public service in all countries, they are publicly provided virtually everywhere.

Public service often also includes sectors like education, health (care and insurance), social insurance, etc. Though not strictly speaking public goods <sup>(1)</sup>, they have

<sup>(1)</sup> Pure public goods are defined on the basis on non-rivalry and non-excludability. Education, for instance, fails the test on both grounds. Exclusion is possible (as illustrated by the existence of private schools) and the benefits to students clearly depend on class sizes.

some specific characteristics which distinguish them from standard private goods which can be effectively supplied in the free market. In particular, it is widely acknowledged that these activities involve strong externalities, have significant redistributive implications and have a crucial impact on the long-run growth perspectives of the economy. Consequently, there is a strong case for public intervention. This somewhat broader notion of public service is widespread in Europe but is less developed in countries like the United States <sup>(1)</sup>.

In many European countries, the notion of public service has traditionally been extended to include public intervention in some of the major network industries (like electricity, telecommunications, postal services, railway transportation, etc.). As for education and healthcare, the rationale for such policies lies essentially in the specific characteristics of the underlying products and services. The goods produced by network industries involve various externalities (including problems of environmental protection). They often constitute essential inputs which are of crucial importance both for national security <sup>(2)</sup> and for the growth perspectives of the economy. They often involve investment decisions that call for long run (and even inter-generational) trade-offs which may not be accurately reflected in available markets <sup>(3)</sup>. Last, but not least, they have strong redistributive implications which lead to the concern that market provision may not be sufficient to guarantee a fair and non-discriminatory access to these services for all income groups and in all locations of the country.

Market failures are of course not specific to network industries, and nearly all goods may have redistributive implications. In addition, public authorities have many alternative instruments (including taxation, and various forms of regulation) which may be quite sufficient to correct (or at least reduce) most market inefficiencies. Put differently, not all market failures call for the establishment of public ownership in the corresponding industry — and this is certainly not what the advocates of public service claim. The case for public

service relies on the argument that in the specific network industries, given the nature and/or the significance of the problems (i.e. the stake which is involved for the society as a whole), more indirect controls are not sufficient. Let us illustrate this argument by two examples drawn from the electricity sector.

First, consider the case of environmental protection. In most contexts, (Pigouvian) taxes constitute an appropriate instrument to correct for the problem of negative externalities. They require less information than direct regulations and they do not interfere more than necessary with the agents decentralised decision-making <sup>(4)</sup>. However, this argument does not go through when ‘very large’ risks (like those associated with the use of nuclear power) are involved. Because of limited liability, the concern that private operators may engage in excessive risk taking appears to be well founded. Consequently, an operator which is under the direct control of the government may appear to be a more appropriate arrangement to warrant an adequate degree of risk taking.

Second, consider the case of investment decisions. Because of the incompleteness of (intertemporal and financial) markets, private firms may have insufficient incentives to engage in investment projects which involve (generally uncertain) returns in the very long run <sup>(5)</sup>. Furthermore, it is not obvious how such incentives could be provided through standard tax and regulatory policy. Consequently, the establishment of a public firm may be the only effective way to warrant the appropriate consideration of long run (and potentially inter-generational) trade-offs.

These two examples illustrate the point that a public firm can potentially perform better than private operators, even if all the traditional instruments of tax and regulatory policy are accounted for. However, it has to be pointed out that the argument relies in a crucial way on the assumption that a public firm can be effectively

<sup>(1)</sup> Even in the USA, public authorities do of course intervene in these sectors. However, the degree of public intervention is significantly less than in most European countries.

<sup>(2)</sup> An economic rationale behind this notion may be the non competitive nature of some markets (the oil market being a prominent example).

<sup>(3)</sup> This is true, in particular, as far as the investment in generation capacities in the electricity sector is concerned.

<sup>(4)</sup> Alternatively, the organisation of markets for pollution rights may prove to be an effective solution which allows for an even lesser degree of public intervention.

<sup>(5)</sup> The research for the development of fusion energy provides a good example. If this form of energy is to be developed, research investments are required now and within the decades to come. The benefits are potentially important (and maybe even spectacular), but they are unlikely to be available in the near future. According to current estimates, fusion may provide a competitive source of energy around the year 2050, and it is not before the year 2010 that the exact prospects of this form of energy will be known.



monitored by the government, which in turn is benevolent. Quite clearly, this is a strong assumption! Certainly, there are agency problems within the public sector, and the benevolence of public decision-makers is by no means self-evident. However, these problems are of a somewhat different nature and their relevance for the design of appropriate industrial and regulatory policies in network industries should not be overstated. Agency problems can be dealt with along the lines of the recommendations drawn from economic theory and the degree of (spontaneous or 'induced') benevolence of public decision-makers crucially depends on the effectiveness of the political process in a democratic society. Any policy measure, ranging from minimalist interventions to the notion of public service can only be effective if policy makers exhibit the appropriate degree of benevolence. Put differently, the case for public service rests on the benevolence of public decision-makers, but so does the justification for any regulatory policy (including the universal service obligation mentioned below).

The natural monopoly property that characterises some segments of these networks (e.g. the distribution and transportation of electricity, mail distribution, railroad and telecommunications infrastructure, etc.) can of course only reinforce the case for the establishment of a public service in these network industries. It has to be pointed out, though, that the underpinning of public service goes well beyond the strict notion of natural monopoly. The natural monopoly in itself, certainly calls for public intervention, but if one neglects all the considerations mentioned above, this intervention can take the rather limited form of a regulatory policy which is applied to otherwise private operators. However, if the society puts a sufficient weight on the other considerations, public service per se, that is the direct intervention of a public operator in the industry, appears to be the appropriate policy.

Quite clearly, this philosophy of public service is hard to reconcile with the liberalisation of network industries. To the extent that a public service requires the direct provision of the good by the public sector, its very nature is challenged by the introduction of competition. However, this does not necessarily mean that all the features of public service have to be given up if an industry is liberalised. Liberalisation can be accompanied by regulatory measures which aim at preserving some of the essential features of public service. The universal service obligation, to which we now turn provides one example of such a regulatory policy.

## **1.2. Universal service**

The universal service obligation imposes a number of constraints on public or private operators which are meant to prevent some of the inefficiencies or redistributional problems that would arise at an unregulated market equilibrium. It will become clear below that its rationale resembles that traditionally used to justify public service. However, it is much more limited in scope. A USO can be an effective way to internalise (network) externalities, to achieve some redistribution between users (of different locations and/or income groups) and to contribute towards the realisation of some public goods (like universal communications networks). However, universal service, in itself, does not address considerations like the long run impact of investment decisions, the security of provision, the environmental impact, etc.

In the remaining part of this chapter, we shall concentrate on the USO and study both the theoretical and practical aspects that this regulatory policy involves. We analyse its rationale, its costs and benefits and the different options available for its practical implementation, both in a monopolistic and in a liberalised environment. We start by assessing the underlying economic issues pertaining to its definition and its rationale (Section 2). This discussion sets the grounds for the remaining, more policy-oriented sections. Next, we deal with the costs (and benefits) of the USO (Section 3), an issue which has drawn most of the attention in the recent literature on the USO. We point out the problems raised by its definition and measurement and suggest possible solutions.

Finally, we study the practical implementation and the financing of the USO (Section 4). We consider and compare several alternative arrangements in various types of environments and analyse their respective advantages and disadvantages.

At this stage, it is important to recall that the sectors covered by this report, though all consisting of network industries, differ in many significant respects. In most of our arguments we try to abstract from these differences by considering some generic form of a network industry. This approach allows us to focus on the main issues and to provide an integrated study of the various sectors. It has, however, the drawback that our arguments may fail to address some crucial specificities of particular sectors. Consequently, some fine-tuning may be necessary to apply the arguments presented in our

study to a specific sector. Even though the number of industries covered by this study prevents us from addressing the specificities of every single sector, we shall devote some attention to the specific features of the various sectors below; see the general conclusions.

## 2. Fundamentals

This section addresses the underlying economic issues. It analyses the economic content of the definition of the USO and attempts to cast this policy within a more general regulatory framework. What type of constraints does the USO effectively impose on an operator? How is it related to other regulatory policies? What are its possible justifications, both on normative and on positive grounds?

### 2.1. Definition and economic content

Our above review of the major network industries has shown that the precise definition of the USO is, to a large extent, country and sector specific. Some crucial features are, however, omnipresent and lead to a generic definition of the USO. From this perspective, the USO can essentially be viewed as the obligation of an operator to offer either a full range or a basic package of services:

- of 'good quality';
- to all users;
- at 'affordable' rates.

The precise definition of the goods and service which are subject to the USO is, of course, sector specific. It also varies across countries, but a number of directives of the European Commission has opened the door for an harmonisation within the EU. Let us provide a few illustrative examples. According to the 1996 Telecommunications act, the universal service basket in the USA includes: voice grade access to the Public Switched Network with the ability to place and receive calls, touch-tone signalling, single-party service, access to emergency services, access to operator services, access to interchange <sup>(1)</sup> services and access to directory assistance. In the UK these service include: basic telephony, message forwarding, directory and operator

assistance, emergency services, phone book provision and the availability of public phones. In the context of postal services, on the other hand, a recent directive of the European Commission stipulates that universal service should at least incorporate the following facilities: the clearance, transport sorting and distribution of postal items up to 2 kg, and the clearance, transport, sorting and distribution of postal packages up to 10 kg and the services for registered and insured items.

In many instances, uniform pricing is imposed as additional requirement. The operator is then not allowed to differentiate its prices (or pricing policies) geographically and/or between consumer types (like households and firms). Whatever its precise definition, the USO is, in essence, a set of restrictions on the operator(s) pricing policy. The requirement to offer service to all individuals imposes a binding constraint only because of the simultaneously imposed restrictions on the pricing policy. If the operator were free to set its prices, the USO would be an empty condition. The operator could then charge any consumer group a sufficiently high price to either cover costs or to ensure that their demand drops to zero. This is no longer true if prices are restricted to be 'affordable' and/or uniform (across consumer types). In this case, prices are likely to be below cost for some consumer types, and the USO constitutes a binding constraint. Put differently, the USO can be seen as a specific form of price regulation; that is, a mechanism through which a public authority imposes prices or pricing restrictions. From this perspective the 'universality' of the service simply means that the pricing restrictions apply to all consumers.

To make the USO operational is not sufficient to define the bundle of goods and services it covers. It is also necessary to make the requirement of affordable rates somewhat more precise. Unfortunately, the translation of this legal (and philosophical) principle into economically meaningful policy guidelines, and from there into precise regulatory measures, is a rather intricate problem. In the telecommunications sector, for instance, the penetration rate can be used as a indicator of affordability <sup>(2)</sup>. However, it certainly falls short of providing a comprehensive assessment. While a declining penetration rate in a given area may well point to unaffordable rates, a constant or even increasing penetration rate does not necessarily mean that rates are affordable.

<sup>(1)</sup> Long distance.

<sup>(2)</sup> See FCC (1996).



Phone service may well be perceived by households as a necessity, and some (even low income) households may remain connected, even though this imposes a significant burden on their budget. This argument is even more compelling in other sectors like electricity for instance. Consequently, it is necessary to assess the burden that the particular service imposes on the budget of specific consumer groups and to determine whether that burden is deemed acceptable by public authorities.

Our theoretical analysis below will shed some further light on the economic implications of the notion of affordability. By analysing the USO as a regulatory pricing policy, we will point to the fact that this problem has to be dealt with in an integrated way. Put differently, the appropriate design of the USO policy cannot rely on a sequential discussion of the different problem involved. Content, price levels and financing mechanisms have to be decided upon simultaneously, and the different decisions are largely interdependent. The crucial question is then to know what is the basic objective of the policy and how can this objective be pursued in the most effective way through the appropriate design of the various components of the Universal Service policy.

Finally, it has to be pointed out that quality introduces an additional dimension of complexity: pricing restrictions are in general complemented by requirements on the operator's quality of service (ranging from 'minimum quality standards' to a precise definition of a range of 'basic' services). Such restrictions can, on the one hand, be an expression of the regulator's concern for quality. Put differently, a high quality may, in itself, be one of the regulator's objectives. On the other hand, they can simply be imposed to avoid the possibility that the operator 'cheats' on the price constraints through low quality. For example, without such restrictions, an operator could effectively bypass a uniform pricing constraint by offering a service of low quality to some consumer groups. This in turn would make the USO an essentially meaningless policy. In the remainder of this chapter, we shall focus on pricing policies keeping, however, quality issues in mind.

## **2.2. Rationale**

We shall now study the possible justifications for a USO. Two different but complementary questions arise. The first one is to know if and how the USO can be justified on welfare grounds, taking into account the various constraints that the policy makers (or regulators)

may face. In particular, one wants to know if the USO can be considered as an effective policy tool, which ought to be included in the public authorities' optimal policy mix. Alternatively, one can adopt a positive approach and attempt to explain why the USO is effectively imposed in most network industries. Acknowledging the fact that social welfare may not be the only relevant consideration for policy makers and regulators, one may explore alternative justifications for the observed prevalence of a USO in network industries.

We review and discuss the major arguments which arise from both of these approaches. We show that it is indeed possible, under plausible conditions, to make a case for a USO based on welfare considerations. While some of the justifications which are often quoted in the literature appear to be either flawed or of limited relevance, we point towards some important aspects which are often neglected. We also show how the political process in itself can lead to the imposition of a USO, irrespective of its implications for overall welfare. Here are the main arguments that have been suggested <sup>(1)</sup>.

### *2.2.1. Network externalities*

As explained in the previous chapters of this report, network externalities arise when the benefits from using a network depend on the number of individuals who are connected to the network <sup>(2)</sup>. For instance, in the case of telecommunications, the number of subscribers determines the number of individuals any particular user can communicate with. Consequently, any individual's decision to subscribe or not to subscribe directly affects the utility of other individuals. However, when deciding upon participation, any particular consumer will only take his own (private) benefits into account <sup>(3)</sup>.

It is often argued that such externalities may lead to an inefficient outcome in an unregulated market. Specifically, one expects that participation will be 'too low', or from a dynamic perspective, that the development of the network will be adversely affected. These

<sup>(1)</sup> See, for example, Cave (1996), Cremer, De Ryckel and Grimaud (1997), Curien and Dognin (1995), FCC (1996), Golay and Dobbs (1996).

<sup>(2)</sup> Networks can also create other types of externalities (e.g. a universal communications network can present the character of a 'public good'). This aspect shall be discussed in Subsection 2.2.3.

<sup>(3)</sup> See Part II, Chapter 1, for more details on network externalities and their implications for efficiency.

inefficiencies may be eliminated, or at least reduced, through regulatory measures which aim at providing access to the network at subsidised rates — and this is precisely what a USO can accomplish.

This points to an often quoted justification of the USO, namely, as a device to correct market inefficiencies in the presence of network externalities. This argument has some appeal for communications networks (telecommunications, postal services, etc.) especially in less industrialised countries, where the networks are in the process of being developed. However, it also has a number of limitations.

- First, it does not apply to all the industries where a USO is imposed. For instance, network externalities can hardly be used to justify a USO in the electricity, gas or water sectors.
- Second, even in those sectors where network externalities do arise, a number of regulatory measures, which are usually associated with the USO, do not appear to have a direct link to this phenomenon. Examples include the uniform pricing requirement imposed on postal operators or their obligation to maintain post offices in rural areas.
- Third, under closer scrutiny, it may not be self-evident that network externalities result in an inefficiently low degree of network participation. For instance, an operator may well find it profitable to 'coordinate' consumers even in the absence of a regulatory obligation. This is because the firm also benefits from the network externalities (they increase the consumers' willingness to pay).

To sum up, the argument based on network externalities, though quite prominent in the debate, appears to provide only a very limited degree of support for the USO.

### 2.2.2. Redistribution

The USO can be seen as a special case of redistributive pricing, that is a policy meant to effect redistribution through prices instead of (or in addition to) income taxation and/or 'direct' transfers. From that perspective it bears some similarities with policies involving 'public provision of private goods', in-kind transfers, etc. The basic feature of these policies is that some essentially private goods like education, child care or health care

are provided either free of charges or at (sometimes highly) subsidised prices.

The recent economic literature has shown that such policies can be optimal in a second-best setting; that is when the policy-makers do not have the necessary information to implement (potentially) more efficient policies like direct transfers <sup>(1)</sup>. A detailed survey of this literature would go beyond the scope of this report and we shall restrict ourselves to reviewing the arguments which are most relevant for the problem under investigation.

The precise rationale of policies like public education or subsidised health care has for a long time been a puzzle to economists. Even though they may create some externalities, education and health care are not, strictly speaking, public goods. In particular, exclusion is usually possible and the marginal cost of serving an additional individual is generally not equal to zero (or negligible). Consequently, one may wonder why the government would find it beneficial to intervene in their provision.

Probably the most convincing argument is that public education or subsidised health care may be a way to reduce some of the most striking inequalities in a society. However, they are not the only conceivable instruments to achieve this goal and to complete the argument, one has to establish their effectiveness with regard to alternative instruments.

Consider the case of health care. Provision at subsidised prices may create over-consumption and thus imply an inefficient outcome. Consequently, if the objective of the public authorities is to help the less-healthy individuals it would appear more efficient to do so directly, through a personalised transfer. Now, this is certainly true in a (hypothetical) world where public authorities can perfectly observe individual characteristics and can distinguish the needy from the well-off individuals. In reality, this is hardly the case and direct transfers to less-healthy individuals may be difficult to implement. Specifically, if sick individuals are entitled to some transfer, everyone has an incentive to pretend that he is needy and the verification of these claims would be impossible or very costly. However, if instead health

<sup>(1)</sup> See, for example, Boadway and Marchand (1995), Cremer and Gahvari (1997) and Guesnerie and Roberts (1984).

care expenses are subsidised, the redistribution appears to be better targeted, even though it may come at the expense of some inefficiency (over-consumption).

To sum up, though potential generators of inefficiencies, price subsidies (or public provision at free or highly subsidised rates) may be an effective instrument of redistributive policies if alternative instruments (like personalised transfers) are not feasible for informational (or other) reasons.

The arguments presented so far can justify various kinds of public policies, ranging from direct intervention (through the creation of a public service) <sup>(1)</sup> to more indirect forms of price regulations. The USO, falls into the second category and it can be used to achieve two types of redistribution.

- Towards high-cost customers (e.g., rural households in the postal or telecommunications sector). This is, for instance, achieved through uniform pricing. However, uniform pricing is not necessary: this type of redistribution occurs whenever price differentials (between consumer groups) fall short of reflecting cost differentials.
- Towards low-income (or otherwise needy) individuals. Prominent examples of measures aimed at effecting this type of redistribution include 'social tariffs' in telecommunications or electricity.

A recent document by the Federal Communication Commission (see FCC (1996)) provides interesting illustrations for both types of redistribution. It explicitly distinguishes 'high cost support' (subsidisation of consumers in 'high cost areas') from 'support for low-income consumers' (which, as is emphasised is not limited to specific geographic areas). Specifically, as far as the second aspect is concerned, it advocates some modifications (and extensions) of the existing 'Lifeline' and 'Link Up' programme. Quite interestingly, this document also recommends universal service support for institutions like schools and libraries (irrespective of their location) <sup>(2)</sup>. Here the redistributive character of the policy appears indirectly, though the subsidisation

of other (essentially) redistributive programs like public education <sup>(3)</sup>.

So far, we have drawn our examples from telecommunications and electricity, the sectors where the redistributive dimension of the USO is the most apparent. In the postal sector, on the other hand, the redistributive argument is no less important, but its relevance less obvious at first. In this sector, cost differentials arise mainly because of the locations (and types) of the addressees; the paying customers, on the other hand, are the senders. To establish that a USO can operate redistribution between, say, urban and rural individuals, it thus has to be shown that the eventual beneficiaries of the USO are not so much the senders, but rather the addressees, and specifically, the high-cost households within this group. This claim is supported by at least three arguments.

First, the very existence of cost differentials according to the location of the addressee is, to a large degree, due to the USO. It is because operators face the obligation of delivering to the addressees mailbox at a given frequency that rural delivery is more expensive. In the absence of a USO, reduced frequency, post office box delivery or similar measures could be implemented by the operator in order to eliminate the excess costs in rural delivery. Such an adjustment would certainly have a much more significant impact on rural households than on the senders of the various mail items.

Second, in the absence of a USO, the postal operators could charge rural households for the delivery cost differentials by imposing a (periodic) fixed fee on those who opt for home delivery (rather than at some collective delivery point) <sup>(4)</sup>. In reality, such 'connecting charges' do, of course, not exist in the postal sector (at least not in Europe), but this does not mean that the operators might not find such a pricing scheme optimal if the USO were removed. Now, the arguments which oppose such differential fees (and which may make them hard to accept on political grounds) are essentially of redistributive nature. From that perspective, the redistributive role of the USO is that it precludes certain (non linear) pricing schemes which would impose a heavier burden on high cost customers.

<sup>(1)</sup> See Section 1.

<sup>(2)</sup> Support for healthcare providers is also advocated, but it restricted to those serving rural areas.

<sup>(3)</sup> See our argument on education and the general problem of 'public provision of private goods' above. Public libraries have rather similar characteristics.

<sup>(4)</sup> The period (as opposed to item based) nature of such a fee should be pointed out. It would thus not violate the traditional principle that the sender pays for the mail item.

Third, a large proportion of letters (and mail items in general) are sent by businesses, and economic analysis shows that firms generally manage to shift (at least part of) costs to their clients. Put differently, cost increases will, in general, lead to price increases <sup>(1)</sup>. Under uniform pricing in the postal sector, banks for instance, have no reason to charge rural customers more for the mailing of their bank statements than they charge their urban customers. However, if mailing costs were different, banks may find it profitable to differentiate fees according to the location of a customer <sup>(2)</sup>. A similar argument goes through for many other types of businesses and, in particular, for mail-order corporations. Consequently, it is very likely, that the burden of a removal of the uniform pricing requirement would eventually fall on the high-cost customers.

To sum up, these arguments have shown that the USO does indeed benefit rural households (implying high delivery costs), so that the first type of redistribution (low to high cost) is certainly as relevant in the postal sector as it is in other network industries. The relationship between USO and income based redistribution (the second type we referred to above), on the other hand, is probably weaker in the postal sector than in telecommunications and electricity. However, one can certainly think as the universal availability of free mail delivery as of an in-kind transfer which, as explained above, can be an integrated part of a redistributive policy.

This role as an instrument of redistributive policy provides probably the most compelling theoretical justification for the USO, at least on normative grounds. The other arguments below provide additional support, but they appear to complement the current argument rather than, by themselves, making a convincing case for the USO.

The economic literature reviewed then suggests two different questions. The first one is to know if it is optimal to use the USO rather than other more 'standard' redistributive instruments (transfers and, say, income taxes). A detailed theoretical investigation of this issue would be beyond the scope of this report (see, for example, Cremer and Gahvari (1995, 1996) for a formal analysis and a review of the relevant literature). We shall therefore restrict our attention to the practical aspects by pre-

senting a simple empirical test which allows one to evaluate the costs and benefits of USO compared to an alternative policy of direct transfer. The second question concerns the optimal design (and financing) of a USO conditional on the fact that public authorities have decided to use this instrument. This is the problem we shall focus on in the third part of this chapter.

### 2.2.3. *Public/merit good*

It can be argued that a uniform and universal communications or transportation network (post, telecommunications, railroad) presents the character of a 'public good' because:

- it 'binds the nation together';
- it is essential for the functioning of a democracy;
- for ethical reasons, society finds it unacceptable that anyone be excluded from communications services.

This argument relies on the idea that the existence of the network is valuable in itself, independently of the specific services it provides to the consumers. Consequently, it may apply even when the provided service is essentially a private good.

The USO can then simply be seen as a way of contributing to the provision of this public good. As mentioned earlier, this argument can be combined with the previous one and it can then explain why redistributive pricing ought to be used in network industries rather than in other sectors where the public good aspect may not be present.

### 2.2.4. *Others*

#### 2.2.4.1. *Regional policy*

The USO can also be an instrument of regional policies. For instance, uniform pricing can be a way to subsidise rural customers, in order to encourage households and firms to locate in rural areas (or to prevent them from moving away). Similarly, maintaining basic public services (like post offices or public phones) in small villages may contribute toward preventing the decline of rural areas.

Though quite compelling at first, this argument has to be qualified under closer scrutiny. The main flaw is that the relationship between universal access to some net-

<sup>(1)</sup> The precise extend of shifting depends on the characteristics of demand and supply and on the market structure.

<sup>(2)</sup> A bank who would not do this would be subject to a competitive disadvantage on the low-cost segments.

works and regional development may be quite complex. Consequently, there may be unwanted side-effects and an overall positive impact is not always guaranteed. For instance, the experience has shown that the access to an efficient transportation network may speed up a region's, decline instead of fostering its development.

#### 2.2.4.2. Partisan politics and pressure groups

So far our approach has been essentially normative. We have studied how a USO can be justified on welfare grounds (taking into account the various constraints that the policy-makers may face). If the main concern of policy-makers is effectively to maximise welfare, these arguments also have a positive bearing and can explain why a USO is imposed in many network industries.

In reality, however, these policies may also emerge for different reasons, associated with the political process itself. For instance, 'rural' pressure groups may advocate uniform pricing because alternative policies (such as direct transfers) are not considered as credible or because uniform pricing is less visible, and thus more easily accepted by 'public opinion'. Similarly, the existence and the scope of the USO could also be explained by regulatory capture. This would be the case if the entrants successfully lobby in favour of strict restrictions on the historic operator's pricing policy with the intent of weakening its competitive position. At the opposite extreme, one can also think of situations where the historic operator itself may use its leverage on the regulator to maintain a stringent USO as this may justify some of its privileges (e.g. monopoly protection in some market segments).

### 3. Cost of universal service

Much of the debate on the USO has concentrated on measuring its cost. However, it appears that the very notion of cost of USO is rather ambiguous; both definition and measurement are problematic. In addition, it will appear that the cost of the USO (whatever its definition) depends on the rest of the regulatory structure; see also Section 4. Consequently, there does not appear to exist a general way of measuring the cost of the USO. Instead, there are several competing concepts and the precise measure that must be used depends both on the question one wants to address and on the regulatory environment.

#### 3.1. Definition

There are at least two possible definitions, or types of definition, depending on the perspective which is adopted.

##### 3.1.1. Profitability cost

The profitability cost can be defined as the loss in profits incurred by the operator due to the USO. Put differently, it measures the 'burden' that the USO imposes on the operator. The proper way to measure this cost consists of comparing the profits realised by the operator at the market equilibria with and without USO. This is obviously not an easy task for it requires estimating the hypothetical equilibrium that would occur if the USO were removed, allowing all the operators to adjust their prices accordingly.

The existing measures (like the 'net-avoided cost' approach) fall short of assessing the so-defined profitability cost. They are essentially based on accounting arguments and they coincide with our definition only if prices and market structure do not change when the USO is abandoned and if the operator has no direct benefits from serving certain 'non-profitable' consumers (reputation, long term strategy, etc.). They can nevertheless be useful as first approximations. In addition, they may have interesting interpretations in some specific contexts. For instance if the USO is financed through cross-subsidies (see Section 4) the methods based on Fully Distributed Cost essentially measure the total amount of cross-subsidies which occur. This estimation may be of some interest but one has to keep in mind that it does not reflect a cost per se.

Note that if the operator is a regulated firm which faces a binding profit constraint (at whatever level), the profitability cost is, in principle, equal to zero <sup>(1)</sup>. Though surprising at first, this simply means that the profitability cost is not the appropriate concept to use in this context. The welfare cost defined below appears to be a more pertinent measure in this case.

<sup>(1)</sup> As long as the USO does not make it impossible to meet its budget constraint; see Gallet and Toledano (1997) for a discussion of this point. The same argument goes through (in the short run) for an operator subject to rate of return regulation. In the long run (when the capital stock is variable) the profitability cost under rate of return regulation is, however, not in general zero anymore.



### 3.1.2. Welfare cost

The welfare cost can be defined as the dead-weight loss implied by the USO. To keep the argument as simple as possible we shall assume here that the dead-weight loss can be approximated by the loss in total (consumer plus producer) surplus. The welfare cost is then obtained by comparing the total surplus achieved at a hypothetical equilibrium without USO to the total surplus realised under the USO.

Let us illustrate this by the simplest possible example. Consider a case of a single operator who faces two types of consumers, high-cost (say rural) and low-cost (say urban) customers. The highest value of total surplus <sup>(1)</sup> is of course achieved if each type of consumer pays a price which equals his marginal cost. Clearly, this implies that rural customers pay a higher price than urban customers. Next, to introduce universal service, assume that the operator faces a uniform pricing constraint. The price will then be some (weighted) average of the respective costs. Rural customers benefit as they now pay a price below cost, while the urban customers lose and pay a price above cost. It can then easily be shown that the decrease in the surplus urban customers necessarily exceeds the increase in the surplus of the rural customers. Put differently, total surplus decreases as uniform pricing is imposed.

One can cast this discussion within the traditional equity-efficiency trade-off framework. Redistributive policies which act through the price system (and lead to 'distorted' prices), have an efficiency cost. This cost has to be balanced against their redistributive benefits which depend on the weights of the different consumer groups in the public authority's welfare function <sup>(2)</sup>. Because of these benefits, the overall welfare-impact of the policy may well be positive. Consequently, focusing on the cost of the USO may be somewhat misleading; the cost is only part of the story and even if it can be

properly defined and correctly measured it does not indicate the overall impact of the policy. A first attempt to correct this problem is presented in the next section.

### 3.2. A broader view: evaluating costs and benefits of universal service

The welfare benefits of the USO (through its redistributive impact) may be even harder to evaluate than its cost. They depend on the weights which the policy makers attach to the different groups of consumers and these weights are, in general, not observable.

Cremer, Grimaud and De Ryck (1997) suggest a simple and operational method, inspired by cost-benefit analysis, which allows one to measure the overall welfare impact of the USO even if the objective function of the policy makers is not known. The idea is to compare the USO to alternative (second-best) policies while holding the redistributive effort constant. Instead of attempting to directly assess the redistributive benefits of the policy they thus use an indirect approach. It consists in comparing two policies (the USO and an alternative instrument) which achieve a given 'amount of redistribution' in order to assess which of them involves the least efficiency cost. The considered alternative policy is that of direct transfers financed through the general budget and involving some 'cost of public funds'.

To illustrate this method, let us assume that there are just two consumer groups, respectively indexed  $r$  and  $u$  (say rural and urban households) <sup>(3)</sup>. Assume that the currently imposed USO benefits  $r$  type consumers, for instance because prices are uniform even though costs are higher for these customers. Next consider the (hypothetical) equilibrium that obtains if the USO is removed. Let  $\Delta U_r$  denote the difference in the surplus of  $r$  type consumers between the USO equilibrium and the equilibrium without USO. Per our assumption that  $r$ 's are the beneficiaries of the USO, one has  $\Delta U_r > 0$ . Similarly, let  $\Delta U_u$ ,  $\Delta \pi^m$  and  $\Delta \pi^c$  respectively denote the corresponding variations in the surplus of type  $u$  consumers, the profits of the USO operator (indexed by  $m$ ) and the profits of the competitors (index  $c$ ). Note that  $\Delta U_u < 0$  while the sign of the other variations is a priori ambiguous. Finally, consider a direct transfer to  $r$  type individuals, implying an efficiency cost of  $\lambda$  per unit (the so-called 'marginal cost of public funds')

<sup>(1)</sup> That is the unweighted sum of the surplus of the producer and that of the different consumer groups.

<sup>(2)</sup> To be more precise, a regulator who is concerned with efficiency only (and whose objective can thus be represented as the maximisation of total surplus, with all consumer groups receiving equal weights) would favour marginal cost pricing. Redistributive objectives, on the other hand, can be introduced by considering an objective function which puts higher weights on some consumer groups. In that case, marginal cost pricing continues to be efficient, but it may not be the welfare maximising solutions. If redistributive objectives are accounted for, it may well be desirable to deviate from the efficient solution in order to switch to a pricing system which is more favourable to consumer groups who receive a higher weight in the regulators objective function (e.g. low income or rural households).

<sup>(3)</sup> See Section 4 for more details on this two group specification.

which is determined to exactly compensate  $r$  consumers for the removal of the USO. Put differently, the redistributive effort is held constant. Cremer, De Rycke and Grimaud (1997) show that the difference between the level of welfare achieved with the USO and that realised under the alternative cash transfer policy (without USO),  $\Delta W$  can be expressed as follows:

$$\Delta W = (1+\lambda)\Delta U_r + \Delta U_u + (1+\lambda)\Delta \pi^m + \Delta \pi^c.$$

This expression provides a simple and operational test for the relative efficiency of the USO compared to the alternative system of cash transfers. If  $\Delta W$  is positive, then the USO is a more effective instrument of redistributive policy than the direct transfers. Intuitively this means that the welfare cost associated with distorted prices is less than that associated with the financing of cash transfers through the general budget. If, on the other hand,  $\Delta W$  is negative, the conclusions are reversed and cash transfers are welfare superior <sup>(1)</sup>.

As far as data requirement is concerned, this test is not more demanding than the assessment of the welfare cost of the USO discussed in the previous section, with the sole exception that it requires an estimate of the cost of public funds  $\lambda$ . Such are available in the literature and are in the range of (0.2–0.3).

Finally notice that the comparison presented here rests on the assumption that cash transfers are indeed feasible on informational grounds. In other words, the ‘needy’ individuals can be identified in a costless way. As argued above, this may not be the case in reality and this problem has to be kept in mind when interpreting the result of the welfare test. It is of no relevance if the calculated value of  $\Delta W$  is positive, for the USO is then unambiguously the better policy. However, some precautions are necessary when the computed value of  $\Delta W$  turns out to be negative. In that case, one needs to have a closer look at the feasibility of cash transfers in the context of the particular sector. If they are altogether not feasible, the comparison becomes meaningless. On the other hand, if their implementation implies a cost, the welfare evaluation should be corrected accordingly,

and this may well reverse the balance in favour of the USO.

#### **4. Implementation and financing**

In this section, we consider, and attempt to evaluate, the different possibilities to organise and finance a USO. We first study the case of a monopolistic operator and then that of a (partly or totally) liberalised sector. Our aim is to go beyond a mere enumeration of scenarios and of their respective advantages and disadvantages. Instead, we cast the different problems within a unified analytical framework which integrates the building blocks we have established in the previous sections. This approach allows us to provide an in-depth study of the essential underlying issues and (hopefully) to provide some new insights which may clarify the debate on the USO.

The monopoly case is of some interest in itself for it continues to be empirically relevant, at least for the time being. In addition, it is a useful starting point for our analysis. It allows us to introduce and analyse a certain number of problems in the simplest possible way. This sets the ground for the analysis of liberalised industries which gives rise to a large set of new issues.

We shall show that under monopoly, the USO and its financing mechanism create a number of ‘distortions’ which adversely affect overall efficiency. This efficiency loss has to be balanced against the benefits (in terms of redistribution, public good provision etc.) to determine the appropriate extent of the USO <sup>(2)</sup>. Moreover, for a given level of benefits, the design of the policy and of the financing mechanism ought to be such that efficiency losses are as small as possible. Though by no means trivial, this problem is rather standard and resembles in many respects a traditional Ramsey pricing problem.

In the presence of competition, on the other hand, additional distortions may arise. The design of the USO and its financing mechanism may now determine the very nature of competition that can be sustained in the sector. It can affect the viability of existing operators as well as the entry process in the industry. To take full advantage of efficiency gains from potential or actual

<sup>(1)</sup> An example of the empirical application of this test is provided by Cremer, De Rycke and Grimaud (1997). This analysis is based on price and cost data, as well as demand estimates, for the French mail service (La Poste). It results in a positive value of  $\Delta W$  (of about FRF 1 billion), pointing towards a positive welfare impact of universal service (in the considered sector and country).

<sup>(2)</sup> The degree at which qualifying consumers ought to be subsidised.



competition it then becomes important to design the USO and its financing mechanism in a ‘competitively neutral’ way. This is a very complex problem as it implies that the regulatory policy must strike the right balance between two potentially conflicting objectives. On the one hand, competitive neutrality requires that no ‘excessive’ protection ought to be granted to the USO operator for this might interfere with the entry process (and jeopardise the viability of potentially more efficient entrants). On the other hand, if the USO is not compensated in an appropriate way, its viability may be threatened by possibly less-efficient entrants (who may find a niche in the market because of phenomena like ‘cream-skimming’). This imposes a threat to both the USO itself, and to the efficiency of the competitive process in the industry.

Our analysis will show that the design of the financing mechanism is the crucial ingredient for the reconciliation of these potentially conflicting objectives. The choice of the appropriate financing mechanism will involve various trade-offs which are, to a large extent, sector (and country) specific. Consequently, it is not possible to determine a single mechanism which would be appropriate in all sectors (and in all countries). A thorough analysis of the various policies is nevertheless useful in that it allows us to reach a better understanding of the advantages and disadvantages of the available mechanisms, so that policy-makers will be in a position to evaluate the different options on an informed basis.

#### 4.1. Monopolistic sector

If there is a single operator in the industry, there are essentially only two ways to finance universal service: cross-subsidies and transfers from the regulator to the firm. Transfers raise the usual issue of whether or not the operator should be required to balance its budget. From that perspective a transfer to finance the USO is very much like a transfer to finance fixed costs and such transfers are often deemed unacceptable for a variety of reasons.

To keep the presentation as simple as possible we start with the case where transfers are indeed ruled out and concentrate on cross-subsidies, the only financing mechanism which is then left. Transfers will be reintroduced later and we shall allow for a financing scheme combining the two instruments.

Let us thus consider the following highly stylised model of a network industry. There is a single, public (or regu-

lated) operator which is required to balance its budget. Further assume for the time being that costs are linear in output and that there is no fixed cost (the impact of more general technologies will be discussed below). There are different types of consumers who differ in their impact on the operators cost (and possibly in other characteristics like income or preferences). Put differently, the (average and marginal) cost of providing ‘one unit of service’ differs between consumer groups. In case of the telecommunications or electricity sectors one can think, for instance, of rural and urban customers. Similarly, in the postal sector, costs (and especially the costs for mail distribution) depend on the location of the addressee (rural or urban) and differ between types of consumers (households or firms).

Ruling out, for the time being, the possibility of non-linear pricing, the USO corresponds essentially to a pricing policy under which (at least some) high-cost customers pay a price below their cost while some other individuals pay a price higher than their cost. Roughly speaking, high-cost individuals are subsidised by low-cost individuals (recall the budget-balancing assumption).

Uniform pricing fits this definition, but it is just one of the possible cases (an extreme case in some sense). Roughly speaking, any policy where price differentials are smaller than cost differentials can be referred to as universal service.

The determination of optimal prices (and, hence, the optimal design of the USO) is then essentially a Ramsey pricing (or ‘taxation’) problem with heterogeneous individuals and with the possibility that the objective function reflects redistributive concern. Notice that universal service and its financing go hand in hand here. Because of budget balancing, price subsidies to high-cost customers can only be financed through (implicit) ‘taxes’ on low cost customers.

Clearly, the solution to this problem does not, in general, involve uniform pricing (over consumer types for a given product or service). However, uniform pricing can be imposed as an additional constraint (justified by political economy and/or horizontal equity arguments). Note that in the case of a single-product firm, budget and uniform pricing constraints combined completely determine its prices; there is no discretion left and the pricing problem reduces to an accounting exercise. For a multi-product firm, however, this is not the case any-

more. Uniform pricing implies a particular pattern of cross-subsidies for any given product, but cross-subsidies between products (if any) are left to be determined.

These arguments remain valid under more general technologies, but the interpretations are then more complicated. In particular, one can introduce a fixed cost. Universal service (and specifically redistributive) considerations then result in prices which differ from the traditional Ramsey prices (obtained by maximising unweighted total surplus). If the fixed cost is sufficiently large it is possible that all consumers pay a price which exceeds their marginal cost. However, because of the redistributive concern, the high-cost customers pay less than they would if prices were set merely according to efficiency considerations.

So far, we have considered only linear pricing policies <sup>(1)</sup>. In many network industries, and in particular in telecommunications and in the electricity sector, pricing schedules are, however, generally non-linear. There is typically a periodic fixed fee which implies that the per unit charge paid by the consumer depends (and generally decreases) with quantity, even if the pricing schedule is otherwise linear. In many instances, quite sophisticated non-linear pricing schedules are used <sup>(2)</sup>. The availability of such pricing policies does not invalidate the arguments presented in this section. However, it adds more flexibility, both for the design and for the financing of a USO. For instance, in the telecommunications sector, cost differentials are to a large extent explained by locational variations in the cost of providing access to the network — marginal costs (of the communications volume) being very similar. The subsidisation of rural customers may then take the form of a below cost access fee, compensated by an above cost charge on urban customers. It should also be pointed out, that non-linear pricing schedules are a particularly attractive instruments as far as the support of low-income customers is concerned <sup>(3)</sup>.

Whatever the specific intend of the policy, the availability of non-linear pricing is always a plus <sup>(4)</sup>. It tends to

reduce the distortions associated with the financing of USO and it is also an effective tool to target the subsidies in a more effective way <sup>(5)</sup>. Nevertheless, the essential features of the financing mechanism described above remain unaffected. Under monopoly, and in the absence of a transfer from the regulator, a subsidisation of some consumer groups is necessarily paid for by other consumers (be it through a higher linear price or through a higher fixed fee).

Finally, let us reintroduce the possibility of a transfer to the operator. First, notice that for any given level of the transfer, the problem essentially remains the same as above (cross-subsidies without transfer) and the pricing rules do not change. Actual prices (and in particular, the extent by which high-cost customers are subsidised) do, however, depend on the transfer. It can be noted in passing that this illustrates a point made above, namely that universal service per se and the mechanism used to finance it are inter-dependent.

The determination of the optimal transfer is a slightly more complicated problem. All essentially depends on how the transfer itself is financed. If lump-sum taxes were available, such a transfer could be financed without any efficiency loss, it would become the dominant instrument (no surcharges would be levied anymore). Under the more compelling assumption that the financing of the transfer also involves an efficiency loss (the so-called marginal cost of public funds) this result does of course not go through. The optimal financing mechanism is likely to be based on both instruments and it strikes a balance between their respective efficiency costs (marginal dead-weight loss of surcharges versus marginal cost of public funds).

#### **4.2. Liberalised sector**

Many of the arguments presented in the previous section remain valid if there are several competing operators. However, as discussed above, additional questions

<sup>(1)</sup> Under linear pricing, the charge paid by a consumer is proportional to quantity. Put differently, the per unit charge is independent of the consumption level.

<sup>(2)</sup> Including menus of two part tariffs, where the consumer can choose between different optional plans, implying each different levels of fixed fees and variable charges.

<sup>(3)</sup> See Cremer and Gahvari (1996) for a detailed discussion of this aspect.

<sup>(4)</sup> As long as the regulator is benevolent.

<sup>(5)</sup> The redistributive properties of non-linear pricing in the public sector are studied by Cremer and Gahvari (1995); see also Philips (1983) and Sharkey and Sibley (1993). Cremer and Gahvari show that non-linear pricing (implemented for instance through a menu of linear contracts) may be an effective way to extract higher payments from large (high-income) consumers, thereby lowering the payments of small consumers. Observe that such a policy in general implies a 'high' marginal price for small consumers (but a low access fee) and a low marginal price for large consumers (combined with a high access fee). On the applied side, Philips (1983) provides an enlightening discussion of 'social tariffs' (based on the pricing policy in the Belgian electricity sector).

arise in such a context. There are now several alternative ways to organise and to finance the USO. In addition, the introduction of competition now introduces additional sources of distortions which may be associated with the financing mechanism. As in the monopoly setting, there continues to be a welfare loss associated with the fact that some consumers pay prices which are above their cost. And the appropriate design of the financing mechanism has to account for this welfare loss. However, the financing of the USO may now create additional distortions in that it may interfere with the very nature of the market structure which can be sustained in the sector. On the one hand, an inappropriate financing mechanism may be an obstacle to the entry of potentially more efficient operators in the industry. On the other, it may also give rise to the emergence of inefficient entry in that regulatory restrictions may foster the emergence of possibly less efficient operators in some market niches. Consequently, the design of the financing mechanism has to account for its impact on the industry structure. If the entry process is otherwise deemed to be efficient, this implies that the financing of the USO has to be achieved in a 'competitively neutral' way, hence minimising its interference with the market process per se.

We shall distinguish between settings where the USO is imposed on a exogenously determined operator and those where the designation of the universal service operator is part of the mechanism used to implement the policy. Exogenous designation occurs, for instance, if the USO obligation is imposed in an ad hoc way on the historic (public or previously public) operator but not on new entrants. Regulatory settings under which the USO is imposed on all operators do also fall into this category, but they give rise to specific problems which need to be addressed. Alternatively, the operator facing the USO can be endogenously determined e.g. through an auction.

#### *4.2.1. The universal service obligation is imposed on a single, specified operator*

Two sub-cases have to be distinguished:

##### *4.2.1.1. The operator who faces the USO is solely responsible for the financing*

This setting is similar to the monopoly case considered above. Specifically, the USO is financed through cross-subsidies between the customers of the corresponding operator. As above, one can also consider the case

where a direct subsidy from the government contributes towards the budget of the operator.

Though similar, the current setting and the monopoly case are, however, not completely equivalent. Specifically, competition may limit the ability of the operator to finance the USO through cross-subsidies. The surcharges levied on some consumer groups may open the door to cream skimming (by possibly less efficient competitors) which creates additional distortions and may threaten the viability of the operator.

This problems can be limited (though not eliminated) through the definition of a 'reserved sector', that is a set of products (services) or activities (like mail distribution) for which the operator enjoys monopoly protection. Nevertheless, the fundamental problem remains: the 'tax base' (the set of goods on which surcharges can be levied to finance subsidies to some consumer groups) is restricted in an artificial way. In the light of the results obtained in optimal tax theory this is likely to bring about a welfare loss.

This point can be explained in a very simple way. It is well known from standard microeconomic theory that the welfare loss (dead-weight loss) of a tax increases more than proportionally with its per-unit rate. Consequently, the welfare loss-per-unit of tax revenue increases as the tax increases. Now, this implies that for a given total tax revenue, the total welfare loss will be smaller if many goods are taxed at a low rate than if few goods are taxed at a high rate. Put differently, the larger the tax base, the smaller the welfare loss. This argument is traditionally presented in the context of commodity taxation. However, it immediately carries over to surcharges levied as part of a system of cross-subsidies which are in essence just a special case of commodity taxes.

##### *4.2.1.2. All operators contribute*

This essentially amounts to creating a universal service fund, financed through implicit or explicit 'taxes' on the operators who are not subject to a USO. The proceeds of this fund are then used to finance a transfer to (partially) compensate the universal service operator for his obligations. The result is a wider tax base (as opposed to the above situation) which, by the optimal taxation argument presented above leads, at least potentially, to a welfare improvement. In addition, the contributions imposed on the competitors may reduce the threat of cream-skimming. If the 'universal service' taxes are

well designed, a competitor can only capture a market segment if he is more efficient than the incumbent operator. Consequently, the competitive process can work in a more efficient way and the survival of the incumbent (universal service) operator is threatened only if he is less efficient than the other firms.

There are several alternative ways to levy the contributions to the universal service fund.

- Universal service taxes (or fees); for instance, specific taxes levied on the competitors' sales.
- Access surcharges; this option is of course only available if the competing operators have to use (part of) the USO operator's network <sup>(1)</sup>.
- Lump sum entry fees, which can be implemented by selling or auctioning off licenses to operate in the sector (see e.g. the postal sector in Germany).

Let us start by discussing and evaluating the first two of these options. The following arguments can be put forward.

- Universal service taxes and access surcharges are equivalent if there is no possibility of bypassing the network and if the network constitutes an input which has to be used in 'fixed proportions' (it cannot be substituted by other inputs).
- Access surcharges appear to involve less transactions cost than taxes. This is because it is sufficient to increase the access fee which is levied anyway. It is not clear, however, how significant the difference really is, especially if the sales of the competing operators are in any event subject to some form of commodity taxation.
- If bypass or input substitution are possible, access surcharges may induce inefficient bypass and/or production inefficiencies.
- Universal service taxes are more transparent to the consumers; the financing of universal service is clearly separated from other issues (marginal cost of access, financing of the network's fixed costs etc.)

which may affect the determination of the access charge.

Summing up, if both of these options are available, taxes appear to be the better instrument. The third option amounts to a lump-sum tax on operators. It should not result in distorted prices (a sunk entry cost does not affect the pricing decisions of a profit-maximising operator) but it may adversely affect entry. Put differently, from a purely static perspective (for a given number of active operators) it appears to be tempting to resort to this instrument <sup>(2)</sup>. From a dynamic perspective, however, lump sum fees may have a negative effect on welfare as they reduce the number of active operators and prevent entry of otherwise efficient firms.

At this point, a very important remark about the 'incidence' of universal service taxes (or entry surcharges) is in order. We have referred to taxes and access charges as being levied on the operators. However, one should keep in mind that their burden (or at least part of it) will fall eventually on the consumers. Standard results in the tax incidence literature are very insightful in this regard. They show that the exact extent to which the tax is reflected in the consumer prices depends on the market structure and on the characteristics of demand and technology, but not on whom the tax is formally levied. Specifically, whether a tax is formally levied on the operators or on their consumers does not affect the way its burden is eventually split between the agents. Put differently the price paid by consumers at the after-tax equilibrium solely depends on the market fundamentals equilibrium and is independent of purely regulatory or legal definitions.

It should also be mentioned that 'pay or play' type taxes, where the competitor has the option of not paying the tax if he accepts himself the USO are a variant of the policies under investigation in this section. They present two additional features.

- They may have the additional advantage of enhancing efficiency. In particular they can prevent the designated operator to 'inflate' the cost of universal service for otherwise the competitors would opt themselves for the USO. In that sense, a 'pay or

<sup>(1)</sup> Cremer, De Rycke and Grimaud (1995) provide a detailed analysis of this financing mechanism for the case of the postal sector.

<sup>(2)</sup> It has to be pointed out though that in a second-best world, the relative efficiency of different outcomes cannot simply be assessed by counting the number of distortions.

play' system presents some similarities with the franchising policy considered below.

- However, they may impose additional monitoring cost on the regulator who may have to enforce the USO for several operators.

Let us finally turn to the issue of how the level of the universal service taxes (or access surcharges) should be determined. At first, one might be tempted to argue that the tax ought to equal the surcharge imposed (price minus marginal cost) by the universal service operator on its own clients <sup>(1)</sup>. However, on closer scrutiny one realises that this is not, in general, correct. Optimal taxes and surcharges can easily be shown to depend on demand elasticities, technologies and market structure. Roughly speaking taxes and surcharges are equal only if the market is perfectly competitive, all operators have the same technology and the products sold by the universal service operator and those offered by its competitors are perfect substitutes.

#### 4.2.2. *Franchising of the universal service obligation*

So far, we have assumed that the USO is imposed on an exogenously determined operator. This has been traditionally the case in many industries. However, over the last few years, different arrangements have been proposed in the debate and are now being experimented in several countries. The essential feature of these alternative policies is that the designation of the universal service operator becomes itself part of the financing mechanism.

One can think, for instance, of the following mechanism. The regulator defines the USO and then organises an auction. Operators submit a bid stipulating the compensation they require to fulfil this obligation and the franchise is awarded (for a given time period) to the least expensive operator. Note that the auctions may be local; that is, pertaining to the USO in a given geographical area.

The Australian system in the telecommunications sector comes close to such an arrangement <sup>(2)</sup>. However, it differs in one respect, namely, that the USO is, in a first

step, granted to one (or several) operator(s) designated by the Minister. Every year, the Universal Service operator then nominates its 'net cost areas', areas where the USO imposes losses. Based on this report, the regulator Austel, calculates the cost of the USO according to the avoided costs method <sup>(3)</sup>. This cost constitutes the basis for the compensation of the USO, which is financed through levies on all 'participating carriers' <sup>(4)</sup>. The regulator is required to publish the result of its calculations and the other operators can then compete for the USO. Specifically, if an alternative operator can credibly document that it will be able to fulfil the USO at a lower cost, it may become the designated USO operator, thereby being entitled to compensation from the other operators. Put differently, even though it falls short of an full-fledged auction based franchising scheme, the Australian system does make the supply of USO contestable.

Franchising and auction based contracting mechanisms have been discussed in detail in Part one (Chapter II) of this report. The reader is referred to this chapter for a general discussion of franchising, auctioning, and more generally contracting procedures involving regulating authorities and service providers. Here, we shall limit ourselves to discussing the specificities introduced by the USO component of the franchising. From that perspective, the franchising system appears to have a number of attractive features.

- It tends to ensure that the USO is assumed by the most efficient operator at a (close to) minimum cost.
- It allows one to avoid a number of distortions associated with the mechanisms based on cross-subsidies (cream-skimming, inefficient bypass, adverse impact on entry) <sup>(5)</sup>.
- It escapes the transactions costs implied by the levying of a universal service tax.
- It requires less information than the alternative arrangements. In particular it is not necessary to

<sup>(3)</sup> See, for example, Cave, Milne and Scanlan (1994).

<sup>(4)</sup> Determined in proportion to 'interconnect time'.

<sup>(5)</sup> The definition of an area offered to franchising is a delicate problem. If it is too large, it involves a significant amount of heterogeneity. Some types of consumers will suffer from the lack of competition within the area if bidding has been done only in terms of the uniform tariff. If, on the other hand, it is too small, low cost consumers may easily bypass the USO.

<sup>(1)</sup> This argument assumes that costs can be determined which is not a trivial problem.

<sup>(2)</sup> See Cave (1996) for a more detailed presentation of this arrangement.



evaluate (marginal) costs for different consumer types, demand elasticities, etc.

On the other hand, it also presents a number of new problems.

- The regulator's expected payment for the discharge of the obligation will, in general, be lower the larger the number of (non-colluding) bidders. Consequently, the franchising scheme may not be appropriate if the number of expected bidders is small and/or if collusion amongst bidders cannot be ruled out. Whether or not this problem is likely to arise depends to a large degree on the specificities of the industries (technologies, number of potentially active operators, etc.). It also depends on the particular auction which is used; for instance, the specification of a 'reservation price' can be expected to mitigate that problem <sup>(1)</sup>. In addition, the local character of the auctions which tends to reduce an operator's start-up costs may also enhance the number of potential bidders.
- In most cases, the franchisee will have to invest in specific assets to fulfil the USO. This raises the question of how to compensate the firm for these investments, particularly in cases where the concession would not be renewed. If the franchiser cannot credibly commit to an appropriate compensation scheme, the franchisee will be induced to underinvest in the specific assets (anticipating the danger of 'expropriation' at the term of the franchising contract) and significant production inefficiencies may result.
- A related problem is the appropriate evaluation of (sunk) assets of the incumbent that may be used by the franchisee. To the best of our knowledge, neither the economics nor the accounting literature provides the necessary tools to address this problem. The rel-

evance of this problem, once again, crucially depends on the specificities of the industry. It appears less important in a sector like telecommunications where existing infrastructures may have become obsolete and where alternative technologies are available (fibre optics or wireless access). However, even in those cases, the pricing of existing assets is important as it determines the speed of adoption of new technologies. At the other extreme there is, for instance, the railroad sector. If the USO concerns the operation of a certain number of trains per week between towns *A* and *B*, there does not appear to be a reasonable alternative to using the existing rail structure.

- The potential role of local communities and administrations raises an additional set of questions. Consider for instance the case of the postal sector where the USO which is to be auctioned off may involve the operation of a post office in a small village. Should the municipality be allowed to participate in such an auction and, if yes, on what terms? An argument in favour of its participation is that because of economies of scope, the municipal administration may well be the most efficient provider of such a service. However, given the complexities of public accounting systems, it appears difficult to organise such an auction on 'fair' terms.

Finally, it should also be pointed out that franchising different areas will not in general result in uniform pricing. While uniform pricing within a given area can be imposed as part of the franchising contract, it appears to be much more difficult to ensure the uniformity of prices throughout an entire country <sup>(2)</sup>. Consequently, it may not be the appropriate solution when (for reasons alluded to above) public authorities intend to avoid geographical price differentials.

<sup>(1)</sup> There is, however, a commitment problem and the announced reservation price may not be perceived as credible. In that case it may fail to effectively deter collusion.

<sup>(2)</sup> In telecommunication, for instance, franchising the USO would concern mainly high cost (low demand). In urban areas, where demand is sufficiently high, there may be room for several competing operators. Now, the price level can of course be part of the franchising contract, but it is hard to predict what will be the evolution of prices in the competitive areas (and prices may well differ between these areas).

## II. Universal service in practice — some experiences

### 1. Introduction

The concept of universal service has undergone a great deal of revisions by both academics and official bodies since the original suggestion of Theodore Vail, the President of AT&T in 1907, that a single supplier should have the obligation to offer a telephone service to anyone requesting it in any particular geographical area. Although the precise nature of these revisions is specific to each industry, in most cases, they have essentially come as a consequence of the increasing supply of new services made possible by the rapid technological progress and the associated structural changes of the economic environment. This has notably been the case in telecommunications and postal services.

This chapter reports on the way this concept is implemented today throughout the world. The emphasis will be particularly put, however, on some specific countries and sectors. These particular experiences have reached such a high level of maturity that their study would undoubtedly provide us with a valuable source for a forward-looking understanding of the fundamental issues related to the implementation of universal service. Chief among these issues are (i) which specific services are to be included in the universal service package? (ii) by which financial mechanism can one ensure that such services will indeed be offered to consumers?

This chapter is organised in two parts. The first part (Section 2) focuses on issues related to the definition of universal service. More specifically, we present, for various countries and sectors, the way universal service has been formally defined and the main features of its practical content. As mentioned above, we focus attention on some specific experiences. Those experiences provide us with some definitions at a level of precision that allows for useful economic analysis. The second part of the chapter (Section 3) examines, for some selected experiences, the various mechanisms designed

to finance universal service and discusses the way these mechanisms have been implemented.

### 2. The definition of universal service

#### 2.1. Telecommunications

##### *USA*

The general spirit of the recent Telecommunications Act in the United States has substantially affected the content of universal service, although more so in its practical aspects than in its basic principle <sup>(1)</sup>. In its fundamental conception, universal service aims at bringing about the provision of quality telecommunications services at affordable rates to consumers, including low-income consumers, in all regions of the nation, in particular, rural, insular, and high cost areas. Furthermore, rural healthcare providers should have access to telecommunications services at rates comparable to those in urban areas and libraries, and elementary and secondary schools should be able to purchase telecommunications services at discounted rates. Let us now turn to a discussion of some issues related to the implementation of this basic principle.

Before discussing the motivations, as described in the 1996 Act, behind the principle of universal service and the extent to which these motivations are fulfilled in practice, let us outline the specific services which are included in the universal service basket. These are voice grade access to the public switched telephone network

<sup>(1)</sup> 1996 Act, Pub. L. No 104-104, 110 Stat. 56. The 1996 Telecommunications Act is a substantial amendment of the Communications of 1934. The following citation from the 1996 Act pretty much describes its general spirit: 'It replaces the paradigm of government-encouraged monopolies with one in which federal and state governments work in concert to promote efficient competition ... At the same time, the statute directs the Commission and the states to work together to preserve and advance universal service, in ways consistent with the new, competitive paradigm.'



(PSTN) with the ability to place and receive calls, touch-tone signalling, single-party service, access to emergency services, access to operator services, access to interchange (longue distance) services and access to directory assistance.

Four criteria provide guidance in the determination of telecommunications services that should be considered for inclusion into the definition of universal service. These are the extent to which (i) the service is essential to education, public health or public safety (ii) the service has been subscribed to by a substantial majority of residential customers (iii) the service is being deployed in public telecommunications networks by telecommunications carriers and (iv) the service is consistent with the public interest, convenience and necessity. Although all of these criteria should be considered in the process of identifying services that fall into the definition of universal service, the Federal Communications Commission (FCC) can include services that do not necessarily meet all four criteria.

Several concerns have been expressed about this approach to defining universal service. A first is whether or not this definition concerns communications services or is limited to telecommunications services (the difference stemming from the fact that communications services can transform the content of transmitted information). This question is important in view of the rapid technological progress in the sector and the FCC adopted the view that the definition should be restricted to telecommunications services. Related to the dynamics of the industry, a concern has also been raised that the above detailed list of services would somehow 'freeze' universal service in the current technology and the services made available by this technology. The FCC adopted the view that universal service should be an 'evolving level of telecommunications services that the Commission shall establish periodically taking into account advances in telecommunications and information technologies and services'.

Although voice grade access to the PSTN was motivated by the need to ensure that consumers may access (local) areas in which essential public services are located, a concern was raised that subscribers in rural areas often need toll calling in order to reach essential institutions such as schools, health care providers and government offices. Hence, basic interchange service has been included in universal service. Touch-tone service is motivated by the fact that it plays an important

role in allowing users to connect to various voice-mail systems, on-line information services (e.g. community bus schedules) and product-ordering services, although the supply of this type of services varies from one state to another.

The inclusion of single-party, rather than sharing of line, service also reflects the forward-looking motivation in the design of the universal service package. Indeed, single-party service is considered as a prerequisite for Internet access. Even though some telecommunications actors recognise that upgrading multi-party service to single-party service might take some transitional time and involve some costs, the FCC has decided that it should be included, in particular, because it is consistent with the public interest, convenience and necessity. Most importantly, it allows access without delay to emergency services (such as basic 911, used to seek police intervention) which are considered as essential to public safety. Access to operator and directory services, viewed as services that assist consumers in the completion and billing of telephone calls (which are widely deployed and used), are considered as essential in public health and safety emergencies as well.

The FCC supports the view that the services included in the definition of universal service should be limited to those carried on a single connection to a subscriber's principal residence and declines connection to second residences even when those residences are located in high cost areas. Single-connection of businesses in rural, insular and other high cost areas are included in universal service because the FCC finds that they share similar general telecommunications needs with residential subscribers, namely, access for health, safety and employment reasons.

The 1996 Act requires quality services. The FCC recognises that it need not require specific technical standards of quality beyond those already adopted and enforced by State quality rules. Hence, the FCC relies on service quality data collected by the State commissions to check that quality service is acceptable. The 1996 Act also states that quality services should be available at just, reasonable, and affordable rates. The concept of affordability is taken in both its absolute and relative dimensions. Hence, the FCC considers that an evaluation that considers price alone is inadequate. Factors other than rates, such as calling area size, income levels, cost of living, population density, and

other socioeconomic indicators are important factors in examining affordability.

The FCC notes that a relatively high penetration rate suggests, but does not necessarily mean, that rate levels are affordable, while a declining penetration rate might be indicative of unaffordable rates <sup>(1)</sup>. While monitoring of demand (subscribership) constitutes a good tool for evaluating affordability, it does not say much about the hardship imposed by the purchase. The FCC considers that it is appropriate to use per capita income and the cost of living in a local or regional area when determining affordability. Because of the important role of these local factors, the FCC gives primary responsibility to the States in evaluating rate affordability.

As indicated above, access to Internet has been considered by the FCC as a necessity and used as a justification for inclusion of single-party service in universal service. On the other hand, the usage of Internet is not included in the Universal Service definition despite the suggestion of some telecommunications actors that it should be. The FCC considers that '... Internet service does not meet the statutory definition of a telecommunications service ...'. The FCC also predicts that, increasing demand for Internet service will eventually circumvent the need to place toll calls to obtain this service and consumers will simply need to rely on access to the PSTN which is already part of universal service.

In a world of rapidly changing technology, incentives for actors to favour the enlargement of the universal service package might exist, as it will be clear from the examination of the financing of universal service in the next section. The FCC finds, however, that an overly broad definition of universal service might offset the fundamental goal of the 1996 Act, namely, preserving the provision of universal service without hindering efficient competition. But the FCC recognises also that the definition of universal service should evolve and be reconsidered in the future. Indeed, the Commission recommends that it convenes a Board no later than 1 January, 2001, to revisit the definition of universal service on the basis of the available information, in particular, the Commission's collected data.

<sup>(1)</sup> Some caution needs to be taken with this procedure, though, as growth of the penetration rate might well come as a result of strategic behavior of firms.

## UK

At the present time, universal service provision is essentially under the responsibility of British Telecom (BT). It is defined as affordable access to basic telecommunications services of all consumers reasonably requesting it regardless of where they are located. These services include basic telephony, message forwarding, directory and operator assistance, emergency services, phone book provision and the availability of public phones. Mercury, BT's competitor, has no obligation of universal service provision. However, its licence requires that its basic telephony offering cover the whole country <sup>(2)</sup>.

The recent publication of a consultation document by OFTEL (December 1995) has generated a large debate <sup>(3)</sup>. The main objectives are geographic accessibility, affordability and equal opportunities for customers with special needs. As a result, special tariffs which are uniform across the country have been proposed. Also, a high level of service for schools and public libraries have been suggested. The availability of advanced services for schools and libraries has been, however, criticized by other actors such as cable companies, on the basis of discrimination, claiming that no-one else in the system has access to these services. Furthermore, the European Commission has found that this differential treatment is in conflict with European competition rules.

## Germany

Recently, German telecommunications legislation has experienced a wave of reforms. The restructuring (with the aim of liberalisation) of the sector was launched in 1989 with the passing of Postreform I. Deutsche Bundespost, the sole telecommunications provider prior to 1989, had a universal service obligation to maintain a ubiquitous network and provide connection to the network to all households and businesses. Affordability and equal opportunities to all consumers was ensured by uniform tariffs across the countries independently of the cost differences associated with distance.

<sup>(2)</sup> Note that the obligation seems to bear more on cable companies, which are quite active in the local exchange business, than Mercury which has come to specialise mainly in long distance.

<sup>(3)</sup> In February 1997, OFTEL has released a consultative document which analyses in great detail most of the important issues related to universal service. In Section 3, in which we discuss some practical aspects of the financing of universal service, we draw heavily on this document.

After Postreform I was passed, Deutsche Bundespost Telekom, the telecommunications services provider that resulted from the breaking up of Deutsche Bundespost, still kept the obligation of offering the basic telephone service all over the country at uniform tariffs. Furthermore, the following services were to be provided by Telekom under the same terms as the basic telephone service: directory assistance, public phones, emergency services (in particular, in public phones), telex and telegraph.

With the Forthcoming Postreform II in 1993, universal service was constitutionalised by an amendment to the German Basic Law. This article states that the Federal Government is under the obligation to ensure the provision of telecommunications services in an ubiquitous, adequate and sufficient manner. Furthermore, it specifies that this obligation will rest on any successor of Telekom be it public or private. In fact, a more precise form was given to universal service by the 1994 Act on the regulation of telecommunications and postal service. This 1994 Act emphasises the role of regulation as a way to ensure a ubiquitous and modern offering of telecommunications services at affordable prices, equal opportunities between rural and urban areas and the achievement of social benefits through this offering of services.

Following the adoption of the resolutions presented by European Council of Ministers, the Federal Ministry of Posts and Telecommunications (FMPT) published a Draft Telecommunications Law aimed at preparing the full liberalisation of the sector by 1 January 1998. This draft notes that the universal provision of telecommunications services can be achieved through regulation. The services concerned are mainly the ones listed above. Some precision is given as to the affordability. Prices for basic telephony are considered as adequate if they do not exceed the (real) price that a representative household living in a non-urban area (an urban area is considered to have more than 100 000 inhabitants) pays for its average demand. This calculation of prices is aimed at ensuring the status quo level of demand by households.

## **2.2. Postal service**

### *EU*

Universal service in postal services has been given great attention by the European Community for various reasons, in particular, because of the very widely spread

locations of users of these services and the associated cost differentials. For instance, the Green Paper on postal services published by the EEC in 1992 had already examined this issue at length. On 15 December 1997, the Council adopted Directive 97/67/EC <sup>(1)</sup> which outlines the conditions under which universal service is to be provided within the Community. More specifically, the directive establishes some rules that identify the services to be included in universal service, tariffs and quality standards.

The directive asserts that Member States should take the necessary measures to ensure that users are permanently provided with postal services of specified quality, all over the territory and at affordable prices. Member States should ensure that, among others, universal service should at least incorporate the following facilities: the clearance, transport, sorting and distribution of postal items up to 2 kg, the clearance, transport, sorting and distribution of postal packages up to 10 kg, and the services for registered and insured items. The directive indicates, however, that the limit of universal service coverage for packages may be increased to 20 kg. Finally, concerning postal packages arriving from other Member States, the directive states that all EC Member States should ensure the delivery, within their territories, of packages up to 20 kg.

Concerning the quality of services, the directive emphasises standards of routing times, regularity and reliability. The specific nature of these quality standards are to be determined by Member States in the case of national services, and by the European Parliament and the Council in the case of intra-Community cross-border services. The quality objective of the directive is that, within the Community, 85 % of all items should be delivered within three working days and 97 % within five working days. The directive, however, notes that some exemptions to these quality standards might be given in some circumstances that may be justified by the specific infrastructure and geography of the concerned country.

As to tariffs, the directive states that they should be affordable to all users and geared towards costs. Nevertheless, Member States might find it appropriate

<sup>(1)</sup> OJ L 15, 21.1.1998, pp. 14–25.

to apply a uniform tariff throughout their national territory. In all cases, though, the Directive indicates that tariffs must be transparent and non-discriminatory. Concerning the cross-border provision of universal service, the directive suggests to Member States that terminal dues, i.e. the remuneration of the distribution of incoming cross-border mail, should be set according to the cost of processing and distributing this mail and the level of quality. Moreover, these terminal dues should be transparent and non-discriminatory <sup>(1)</sup>.

### 2.3. Railway transport

#### EU

The services provided by railway transport companies might roughly be classified into four categories. Local and regional transport corresponding to a small distance (less than 50 km) and a short duration (less than an hour), intercity transport, international transport and freight transport.

Member States have some impact on the way these services are provided since, typically, a contract is signed between the operator and the State or regional authority. There are, however, some guidelines on the way these contracts should be set up which are described in the Council Regulation (EEC) No 1893/91 <sup>(2)</sup>. This rule specifies the obligation for the operator to ensure continuity and regularity of service and some capacity, and various other obligations such as no-exclusion, affordability of prices and price discounts to particular groups (students and retirees).

There is a strong obligation for the operator to provide local and regional service of regular frequency at a fair level of unit price which is independent of time of day but decreasing in distance. There is a generalised trend towards the softening of the universal service constraints on national service, nearly no constraints on freight transport and no constraints at all on international service.

### 2.4. Water

#### EU

The notion of universal service in the sector of water refers to the obligation for the operators to provide water of good quality to all consumers at affordable rates.

Within the EC, standards of quality are set through European norms. The first of these norms (Council Directive 80/778/EEC) establishes the quality of water used for human consumption. It specifies some allowed concentration ratios for 62 parameters concerning drinking water. The second (Council Directive 91/27/EEC) concerns the processing of used water in urban areas, hence its purification. It requires that communities invest in equipment necessary to collect and purify used water (sewerage system) within a specified time frame.

These norms are rather severe and only a few countries succeed in meeting them. Some countries, such as Portugal, Greece and Finland, are lagging behind and thus will have to invest heavily in technology in order to comply to these norms. No individual country in Europe has a stronger regulation than that dictated by the EC norms which, therefore, are binding for all European countries. The objective is to provide European residents with good quality water, which, in some countries, currently, needs to be improved. It should be noted that the norm concerning the purification of used water is the most difficult to satisfy and, today, precisely, quite a few countries do not meet this norm.

The obligation to serve all consumers bears on both the distribution of water and the purification of used water. Concerning the distribution, the operator has to supply water only to those households that are connected to the network, which constitute the large majority of households. The same obligation applies to the purification of used water, although fewer households are connected to the purification network than to the distribution network <sup>(3)</sup>. The households which are not connected to the network have, in general, their own water-purification facilities.

<sup>(1)</sup> Recent multilateral agreements among countries of the Universal Postal Union (UPU) indicate that operators have the obligation to deliver mail of foreign origin at 'terminal dues' representing 80 % of domestic price. Furthermore, in case of insufficient quality of service, some penalties are imposed on the operator of the country of destination.

<sup>(2)</sup> OJ L 169, 29.6.1991, p.1.

<sup>(3)</sup> On average, in Europe, 89 % of households are connected to a distribution network, whereas only 76.5 % are connected to a sewerage system.

This is, in particular, the case in Finland, Italy, Ireland, Spain and Greece where the households which are not connected to the network of purification of used water are equipped with septic tanks.

The affordability of rates is linked to the requirement of balanced budget of the operators. Prices are set just to recover costs of supplying water. Hence, average cost pricing seems to be the rule, with three notable exceptions though, Italy, England and Wales. In Italy, water is sold at below cost and in England and Wales a price cap system is utilised which is revised annually according to profit levels. In the case where water supply is under local government authority or government-owned firms (e.g. Germany, France, Italy, Finland, Portugal, Sweden, etc.), the price is set by the local authority which ensures affordability. If the network is managed by a private firm, still a State agency exercises some oversight on its pricing policy. For households that do not possess a water meter, a tax determined by the city hall is levied on them.

#### *USA and Japan*

In the United States, while states have some autonomy with respect to the establishment of norms of quality for drinking water and purified water, various rules have been set at the federal level. These rules include the Federal Water Pollution Control Act Amendments (1972), the Clean Water Act (1977) and the Water Resource Development (1986). Nowadays, the legislation which governs the universal service obligation in terms of quality norms is the Drinking Water Regulation.

In Japan, the law on water passed in 1970, specifies two series of norms: one related to the consumption of water by households and the other to the protection of the environment. Following some serious chemical contamination of drinking water in 1990, the Ministry of Health has established a new legislation which incorporates stricter norms of quality.

In both the United States and Japan, operators do not have an obligation to provide drinking water to household located outside their serving area, i.e., those who are not already connected to their network. In both countries, affordable prices are implemented through a balanced-budget constraint imposed on the operator by some local regulatory authority.

## **2.5. Electricity**

Universal service in this sector consists of the obligation of electric utilities to supply service in a continuous manner, to meet the needs of all customers requesting it, and provide it at the minimum possible price without damaging the environment. In no country does there seem to exist an obligation for the power distribution operator to connect individual users to the main network: indeed, constructing dedicated lines and transformers for potential isolated customers would cost more than installing individual generators at the consumption site. However, most countries include, in their national or regional plans, the connection of newly developed residential and industrial areas to the power distribution networks.

The continuity obligation is the guarantee, in time and space, of secure power supply. Roughly speaking, this means that an electricity provider must guarantee the continuity of this provision to any connected customer, unless a contract for load-shedding has been agreed upon. The electric power utility must take the necessary actions to meet total demand and provide service at the minimum cost. This obligation, although not a legal obligation per se, is included in all countries' regulatory schemes. Finally, in all countries the above obligations are implemented under constraints imposed by environmental protection.

In some countries, such as the UK, where the generation of power and its distribution are done by separate entities, there is a particular aspect of service obligation which is worth mentioning. Electricity generation firms are usually bound by a contract to provide some level of supply taking into account their capacity. If the supply level is inadequate, however, the very stability of the network may be called into question. Hence, even though these generating firms do not have a universal service obligation per se, the threat of breakdown of the network by itself puts them in a situation of a de facto obligation of continuous supply within their contractual commitment. The authority in charge of the generation pooling and dispatching has the ultimate obligation to monitor demand and supply. In case of foreseen shortage, it may rely upon alternative domestic generating firms and imports from France or Scotland. In Spain, a new (but not yet implemented) law allows the provision of power supply by an independent industry besides the integrated one. If the power generating firms of the independent industry cannot meet their delivery com-



mitments, the integrated sector is in charge of fulfilling those commitments.

Finally, distribution firms have an explicit obligation to serve their clients.

## 2.6. Gas

There seem to be very few universal service obligations in the gas industry. Indeed, in contrast with telecommunications, postal and electricity services to which almost everyone has access, a non-negligible fraction of populations is not connected to gas network. There exist, however, various substitutes to gas for both residential and industrial services.

## 3. The financing of the universal service

### 3.1. Telecommunications

#### *EU*

Various measures have been, and are being, put in place at the European level in order to prepare for complete liberalisation of the telecommunications markets in most of the countries of the Community by 1 January 1988. It has been recognised, however, that actions need to be taken to ensure that universal service is delivered to residents in the new competitive environment. The Community has established a framework that sets limits on what may be funded from a universal service funding mechanism and gives details on how such a mechanism should operate. The Full Competition Directive states that Member States ought to clear from the Community any plan of implementation of a universal service funding mechanism.

The Commission has set detailed guidelines on the costing of universal service. In particular, the Commission insists on the need to incorporate in the costing methodology the notions of unavoidable costs, efficiency and intangible benefits (although recognised by the Commission to be difficult to evaluate) of being a universal service provider such as enhanced brand recognition, benefits from ubiquity, life cycle and advertising benefits. Within this framework, however, Member States have some discretion over how specific arrangements are to be made, in particular, on whether or not the need for a special funding arrangement is warranted, the scope of the universal service basket, the precise nature of the funding mechanism and the mode of determination of the providers of universal service

and the contributors to its funding <sup>(1)</sup>. The recent experience in the UK (which we describe in some detail below), in which the telecommunications market has reached a great level of maturity, provides us with invaluable information.

#### *UK*

The state of play concerning the funding of universal service in the UK might be summarised by the two following consultative rounds initiated by OFTEL. At the end of 1995, OFTEL has published an initial consultation document in which it proposed a levy on all telecommunications operators to finance a universal service fund. In a recent (February 1997) consultative paper, OFTEL has rejected the idea of a fund to which British Telecom's (BT) competitors would contribute to compensate it for providing universal service. Although OFTEL does not consider that, for the time being, there is a need to establish a funding mechanism, it recognises that such a need might arise in the future. Consequently, it has made some propositions on the kind of mechanisms which might be appropriate and committed itself to review the universal service arrangements in 1999.

The decision has been based on an evaluation of the cost and benefits to BT of providing universal service. The main features of this cost/benefit analysis are reviewed below. Given that some room is left for a possible institution of a universal service funding mechanism in the future, OFTEL has laid down the basic principles that should govern such an arrangement. Some of these guiding principles are discussed below.

In order to determine whether or not the universal service provider (BT) needs to be compensated, OFTEL has analysed both the cost and benefits associated with holding the obligation of providing universal service. OFTEL considers that direct financial costs are only part of the story and, indeed, argues that there exist benefits associated with the obligation, although it recognises that more work needs to be done to evaluate those benefits. Let us examine the cost and benefits sides of OFTEL's argument in turn.

<sup>(1)</sup> It should be emphasised that, as recognised by the Commission, whether or not an arrangement based on a universal service fund is needed depends crucially upon some specific features of each Member State. Factors such as population density and topography have a major impact on the cost structure of the network and, therefore, are expected to be determinant in the decision to set up a universal service fund.



OFTEL considers that there exists a universal service cost (gross from the benefits that will be discussed below), if the operator's revenues from serving a customer or a group of customers do not cover the costs it incurred in providing service to those customers. This universal service 'deficit' might arise because the operator has to apply a uniform tariff both across customers and geographical areas, even though the costs of serving them differ. OFTEL considers that these direct financial costs of universal service should be measured by the difference between foregone revenues and long-run avoidable costs. It particularly draws attention to the fact that revenues should take account of, besides line rentals and connection charges, both incoming and outgoing calls.

The elements of universal service that are costed are 'uneconomic' (unprofitable to the operator) areas, customers and public call boxes. Other components of universal service which are not included in the costing analysis are BT's maritime services, emergency services and services for the disabled; the reasons for not including them being that they are funded from other sources or they are an obligation upon all operators. Based on available data for 1995/96, OFTEL estimates that less than 0.5 % of the total number of UK lines are uneconomic and impose a universal gross cost of GBP 10 million to GBP 15 million. Uneconomic subscribers and public call boxes represent 6 % to 7 % of the total UK subscribers' base and about 20 % of the total number of BT public call boxes, with estimated net cost of GBP 45 million to GBP 55 million and GBP 10 million to GBP 15 million, respectively. Hence, an estimate of net cost to BT of universal service of GBP 65 million to GBP 85 million has been obtained by OFTEL.

OFTEL has argued that the cost of universal service should be adjusted for efficiency of production, the reason being that if such a cost constitutes the basis for determining the contributions of other operators to the funding mechanism, those operators should not be expected to pay for the inefficiency of the universal service provider. A downward adjustment factor of 5 % has been applied to the above total cost (to obtain an estimate of the efficiency level of avoidable costs), based on the estimation that BT's operating costs are 5 % higher than the costs of the most performant LEC in the United States (also, the assumption that the same inefficiency factor applies to capital costs has been made). Given that the efficiency adjustment factor reduces avoidable costs (and that universal service cost is the difference between revenues and avoidable costs),

the impact on cost of universal service can be substantially higher. OFTEL has estimated universal service cost adjusted for efficiency in the range of GBP 45 million to GBP 65 million.

OFTEL has also conducted a study aimed at estimating the cost of universal service in the future. Concerning the servicing of the uneconomic areas, OFTEL considers that, given the technological developments for which BT has access, forward-looking calculation of avoidable costs should reduce those costs by about one-half. For the economic customers, the main change in the universal service cost in the future might come as a result of the introduction of new services into the universal service package. An estimate of the impact of this variation of the universal service basket has led OFTEL to adjust the cost of universal service due to uneconomic customers upward from GBP 30–40 million in 1996/97 to GBP 40–60 million in 1998/99. As to future universal service cost of uneconomic public call boxes, OFTEL has considered that since BT has been given more flexibility to re-site its public call boxes, these costs should be reduced. Overall, OFTEL has come up with an estimate of the total cost of universal service adjusted for efficiency, for 1998/99, that ranges from GBP 45 million to GBP 80 million. Recall that these cost estimates are gross of benefits of providing universal service and that those benefits are considered as important by OFTEL. Let us now outline some of these benefits.

In order to evaluate the net cost of universal service, OFTEL considers that some current or future benefits stemming from the provision of universal service should be subtracted from the above estimated costs. OFTEL realises that quantifying those benefits is a difficult task and one can only hope for a rough estimate <sup>(1)</sup>. The precise nature of these benefits, OFTEL realises, might be different between uneconomic areas and customers, on one hand, and uneconomic public call boxes, on the other hand.

Three types of beneficial effects are identified by OFTEL as possibly related to the servicing of non-eco-

<sup>(1)</sup> OFTEL interestingly notes that 'in principle, the scale of these benefits would be revealed in a competitive auction for the minimum subsidy that an operator would require to take on the responsibility for providing specified elements of universal service obligation...'. Furthermore, 'OFTEL intends to explore the possibility of tenders for parts of universal service...'. (OFTEL, February 1997, 'Universal Telecommunication Services', Consultative Document.)

conomic areas and customers: 'life-cycle effects', 'ubiquity' and 'brand enhancement and corporate reputation'. Life-cycle beneficial effects to an operator servicing uneconomic areas and customers might exist because providing service to those areas and customers now might increase the probability of servicing them later when they become profitable. Because new households to an area might not be aware of the existence of BT's competitors, OFTEL (and BT, although there was a disagreement on the size of the effect) considers that BT obtains an advantage from ubiquity. OFTEL considers that servicing uneconomic areas and customers has the effect of enhancing the brand image and, more generally, the corporate reputation of the service provider. This might translate into beneficial effects on overall current and future profitability, e.g. by slowing down the loss in BT's market share due to competition. OFTEL's estimates led it to conclude that '... the size of the benefits in aggregate is likely to be sufficiently large to offset the estimated universal service costs' of servicing uneconomic areas and customers (see OFTEL, 1997).

Concerning benefits of serving uneconomic call boxes, OFTEL identifies two types of effects: life cycle effects (motivated here by the existence of a significant variability in revenues from individual call boxes over time) and the value of advertising of BT's logo on call boxes and the subsequent positive effect on the corporate reputation. Again, OFTEL concludes that these benefits would certainly offset the universal service cost of public call boxes.

Although OFTEL realises that the results of its attempt to quantify the cost of universal service, net of benefits to the universal service provider, heavily depends upon the quality of the data (which it invites BT to improve by supplying more precise figures), it came to the conclusion that, for the time being, 'there is no proven case that there is an undue financial burden on BT that would justify setting in place new universal service funding arrangements' (see OFTEL, 1997).

We have discussed above how OFTEL has reached the conclusion that there is currently no need to set up funding arrangements to finance universal service. However, OFTEL recognises that this need might arise in the future. Indeed, if an undue cost burden on the universal service provider were proven to exist, OFTEL considers that it would be appropriate to put in place funding arrangements in which all public operators

would contribute to the net cost associated with universal service <sup>(1)</sup>. In the event these arrangements are to be made, some important issues associated with their implementation are explored by OFTEL. Let us say a few words about each of these issues.

First, there is the obvious question of who the contributors to the funding mechanism would be. OFTEL considers that, since universal service concerns society as a whole, if any cost burden were to arise for its provider, it should be spread over as wide a cross-section of operators as possible (that is to say, over the largest possible cross-section of users). Hence, OFTEL's view is that '... all public network operators with an individual Telecoms Act licence could be potential contributors ...'.

Second, what would be the basis for the calculation of contributions? Two directions are explored. One might take the view that contributions ought to be set according to revenues, more specifically, in relation to as wide a revenue base (from telecommunication services) as possible. Alternatively, one might challenge the formidable task of relating the contributions to the benefits that universal service brings to users <sup>(2)</sup>. Although some operators have suggested the use of call minutes as a basis for determining contributions, for reasons related to ease of information collection and auditing, OFTEL has favoured revenues <sup>(3)</sup>.

Third is the issue of how would the funding mechanism be administered. OFTEL has explored two alternatives. One which would require the settling of an actual fund administered by an independent body. Another which wouldn't require the creation of such an independent institution, but rather would rely on the compliance of all concerned operators to some specified rules of organisation of the financing of universal service. OFTEL refers to the latter option as a 'virtual fund'. In both options, OFTEL has the responsibility of specifying

<sup>(1)</sup> Such a funding mechanism is not expected to be set up before the review of the net cost of universal service by OFTEL in 1999.

<sup>(2)</sup> An effect of the universal service obligation is to increase the number of users of the network relative to a situation without universal service obligation. Hence, this creates a positive externality on the profitable subscribers as they can access and be reached by a larger network, in particular, the network of unprofitable (universal service) subscribers. See OFTEL (February 1997) for a further discussion of the benefits of universal service to customers.

<sup>(3)</sup> A further justification is that, broadly speaking, two customers with the same telecommunication bill would make the same contribution to the fund under a revenue-based contribution system, which is not necessarily the case under a usage-based contribution system.

ing the costing methodology to be used, the concerned operators and the basis for the calculation of contributions. Clearly, a virtual fund approach has the feature of being more decentralised than an actual fund approach. Although the actual fund approach has had the support of OFTEL in its December 1995 consultative document, the potential high costs of administering such an actual fund has recently (as of February 1997) led OFTEL to favour the virtual fund approach.

Finally, as competition is leading the way in the telecommunications industry, OFTEL has considered the possibility of using market forces as a means of inciting the provision of universal service in the economy. As the principle that universal service is costly and, hence, 'someone has got to pay for it' is generally well accepted, the introduction of competition in the delivery of universal service would ensure that this is done in the most efficient way. OFTEL has explored two incentive-based mechanisms: the auction and the 'pay or play' mechanisms. A brief discussion of some of the practical issues raised by these two methods is certainly worthwhile.

Competitive tendering for areas could, potentially, be a useful means of testing whether or not there exists a net universal service cost of serving uneconomic areas <sup>(1)</sup>. In practice, the idea is to auction off the universal service responsibility for specific areas that include both potentially profitable and non-profitable sub-areas. This would encourage the most efficient delivery of service in those areas. The same scheme might be used for public call boxes. In practice, the responsibility to provide reasonable access to public call boxes can be merely incorporated in the contract for servicing an area which is tendered. But, strictly speaking, a group of economic and uneconomic call boxes can be subjected to a tender process <sup>(2)</sup>. Because areas might include both profitable and unprofitable customers, tenders can also be used for some residual obligations such as the provision of new services included in the universal service package.

The performance of the tendering mechanism described above depends crucially upon the organisation of a genuinely competitive auction. One difficulty might arise

from the fact that there might be relatively few bidders in any given area, in particular, if it is considered as uneconomic. Indeed, in practice, only the operators that have, or are willing to invest in, costly infrastructure in the area in question might be interested in the tender. The process is therefore vulnerable to strategic behaviour on the part of the bidders. OFTEL has explored some ways of designing the process so that the undesirable effects of market imperfection are minimised.

In order to counter collusive bidding, OFTEL has explored the idea of fixing a reserve price based on its estimation of net cost of universal service to the current provider (BT). If no better tender bid has been made, the status quo is maintained, i.e. BT continues to provide universal service. Also, OFTEL considers that a single-round auction of sealed bids might be preferable to a multi-round auction that might leave some room for collusion. Sequential tenders for different areas have been considered by OFTEL as useful for bidders to learn from the conduct of the earlier tenders, but might well invite collusive behaviour.

OFTEL draws attention to the 'winner's curse' problem which might arise because of the asymmetric information between bidders on the costs of servicing some given areas. Indeed, because of the large demand of desegregated data needed to evaluate the net cost of universal service, the incumbent (BT) might be at an advantage, with respect to its competitors, when formulating the size of its bid. Hence, because of this informational disadvantage, a competitor would only win the tender at a subsidy (for the provision of universal service) insufficient to cover its net costs. Finally, OFTEL has anticipated the situation in which, because of the high entry (infrastructure) cost, no operator competes with BT for a tender as it might likely be the case for uneconomic areas. One possible way considered by OFTEL then is to have BT transfer or lease assets to potential universal service providers. Besides the legal and practical difficulties that needs to be taken care of, this option might put the incumbent (BT) in an advantageous position when bidding against operators relying on transfers or lease of assets from their competitor.

An alternative method of using market incentives for the delivery of universal service that has been explored by OFTEL is the idea of 'pay or play' whereby an operator could choose voluntarily to provide service to uneconomic customers and get in return the eligibility to receive universal service funding. Naturally, in a

<sup>(1)</sup> The critical assumption here is that of a truly competitive auction, which, in practice might be violated, as discussed below.

<sup>(2)</sup> In theory, an auction might be organised for each individual call box. However, this process is most likely to be impracticable.

context where all operators make contributions, this funding would be discounted off the operator's contribution to the net cost of universal service.

### 3.2. Postal service

#### EU

As a general rule, universal service is financed through cross-subsidies combined with the exclusive reservation of some specified markets corresponding to services included in the universal service basket, the so-called 'reserved area', for the universal service provider(s). In addition, Member States have the authority to establish a compensation fund if the cost burden on universal service providers is large.

Article 7 of Directive 97/67/EC <sup>(1)</sup> specifies the services which may be reserved for the operator(s) that provide universal service. Each Member State may take the necessary measures to reserve for the universal service provider(s) the following services: clearance, transport, sorting and delivery of items of domestic correspondence, whether by accelerated delivery or not. These items should satisfy simultaneously two criteria: a price criterion and a weight criterion. Their price should be less than five times the public tariff for an item of correspondence in the first weight step of the swiftest standard category where such a category exists. Their weight should be less than 350g. Moreover, to the extent that it is necessary for the viability of universal service, cross-border mail and direct mail may continue to be a reserved area, within the price and weight limits defined above <sup>(2)</sup>.

The directive also addresses the issue of the safeguard of universal service against 'cream-skimming'. Non-reserved services may fall either within the domain of universal service or outside. For those services belonging to the universal service basket, which, potentially, may pose the problem of cream-skimming in a significant manner, the directive indicates that Member States may introduce authorisation procedures such as individ-

ual licences. Although the problem may be less severe in the case of non-reserved services lying outside universal service, such authorisations may be desirable in order to guarantee compliance with essential requirements. Again, the directive states that the implementation of such measures is left to the Member States' discretion.

As indicated above, Member States may establish a compensation fund (administered by an independent body) to further ensure the financial feasibility of universal service. Indeed, it may well be the case that the cost burden imposed by universal service jeopardises the financial viability of its provider(s). In this case, the granting of authorisations to offer non-reserved service may be made contingent upon actual contribution to that fund. It goes without saying that since the granting of authorisations is left to the Member States' discretion there is a need for a harmonisation of these procedures. The directive indicates that this harmonisation process is on the EU agenda.

The directive states that the universal service providers ought to keep a clear and transparent accounting system. In particular, it emphasises the obligation for universal service providers to keep separate books for reserved and non-reserved services. Furthermore, the accounts for the latter services must specify without any ambiguity the services which are included in the universal service package. Finally, the directive specifies the principles for the allocation of those costs which cannot be directly assigned to a particular service, and the rules to ensure the enforcement of the separable cost accounting system.

#### Germany

As indicated above (in the case of telecommunications), an amendment to the German Basic Law has included in the constitution the obligation of the federal government to ensure provision of universal service. Anticipating the liberalisation of the sector, to be effective as of 1 January 1998, it has been decided that this universal service obligation applies to all potential operators.

Entry into the industry is to be organised through licensing. However, as discussed in the definition of universal service, an exclusive licence is granted to Deutsche Bundespost, the historical monopolistic provider of postal and telecommunications services (divested from its telecommunications activities), for

<sup>(1)</sup> OJ L15, 21.1.1998, pp. 14–25.

<sup>(2)</sup> Liberalisation of the postal market within the EU will proceed in different steps. The directive states that by 1 January 2000, the EU should make decisions towards the liberalisation of cross-border and direct mail. Furthermore, the EU will review the price and weight limits taken account of the economic, social and technological developments of the moment, as well as the financial equilibrium of the universal service provider(s). These decisions are to be effective 1 January, 2003.

the provision of regular letters weighing less than 350 g and the price of which is less than five times the basic rate. This exclusive licence is to be maintained until full liberalisation (January 2003). The monitoring of universal service obligation is then to be carried out by an independent regulatory agency.

If the terms of universal service obligation are not satisfied, the agency can impose compliance on the 'dominant' licensee. If the regulatory body expects a potential deficit related to the provision of universal service, it can organise an auction for the universal service obligation. If this auction is indeed organised, the regulatory body pays the winning operator the amount of the bid. If no bidding process occurs and a deficit turns out to be realised by the universal service provider, then the regulatory body compensates this operator for it. In both cases, the funds come from contributions of all operators whose sales are in excess of DEM 500 000 in proportion to their market shares.

#### *Sweden*

The Swedish postal services market was deregulated in January 1993 by an Act of Parliament which put an end to the historical monopoly of the Post Office, although one should note that, to a very large extent, the Post Office de facto monopoly remains. The Post Office faces universal service obligation and receives compensation for some specific services (such as the distribution of the literature for the blind). A current (particularly heated) debate, however, raises some questions about the competitive behaviour of the Post Office and, therefore, the appropriability of such compensations <sup>(1)</sup>.

### **3.3. Water**

The drinking water industry is essentially monopolistic. In general, each large city is served by a network which

is the property of either a private or public firm. This is the case for the purification of used water as well. In smaller cities and rural areas, typically, the local authorities join their efforts in delegating a private or public firm for the management of a common network.

#### *EU, USA and Japan*

Cross-subsidies going from (urban) low-cost customers to (rural) high-cost customers appear to be the main instrument used to finance universal service in the water sector. More specifically, average cost pricing (corresponding to the balanced budget constraint imposed on the operator) allows the operator to use the profits from serving low cost customers to subsidise the losses from serving the high cost customers. Moreover, in some countries including Germany, Austria, Italy and Ireland, the water supply activity is merged with some other activities such as urban transport, electricity and gas <sup>(2)</sup>. Cross-subsidies may then operate across these activities whereby losses in one activity are compensated by profits in another.

Direct transfers from local or State governments to the operator are an alternative way of financing universal service. This is the case for instance in France and the UK, where, respectively, the 'Agences de l'Eau' and OFWAT and NRA provide the financing of new investments in maintenance of existing equipment. The EC is also involved in the financing of these investments, in particular in less advanced countries (Portugal, Greece, Spain, etc.). Generally speaking, the financing of these investments is shared among various institutions including the city hall (usually the owner of the network), the managing firm (if applicable), local government bodies (conseils généraux in France, *Länder* in Germany, provinces in Spain, states in the USA) and the central government (Federal Government in the USA).

<sup>(1)</sup> See E. Nerep (1996) for a discussion of this debate.

<sup>(2)</sup> In France, such (financial) merging of activities is prohibited ('Norme Comptable M49'), although water companies may offer other local services.



# References

- Boadway, R. and Marchand, M. (1995), 'The use of public expenditures for redistributive purposes', *Oxford Economics Papers*, 47, pp. 45–59.
- Cave, M. (1996), 'Meeting universal service obligations in competitive telecommunications markets: lessons for the postal sector' in *Cost of Universal Service*, edited by U. Stumpf and W. Elsenbast, Bad Honnef, Germany: WIK Proceedings, pp. 95–114.
- Cave, M., Milne, C. and Scanlan, M. (1994), 'Meeting universal service obligation in a competitive telecommunications sector', *European Commission*, DGIV, Brussels.
- Cremer, H. and Gahvari, G. (1995), 'Non-linear pricing and redistribution', GREMAQ discussion paper.
- Cremer, H. and Gahvari, F. (1996), 'In-kind transfers, self-selection and optimal tax policy', *European Economic Review*, 41, pp. 97–114.
- Cremer, H., De Rycke, M. and Grimaud, A. (1997), 'Cost and benefits of universal service obligations in the postal sector' in *Managing Change in the Postal and Delivery Industries*, edited by M. A. Crew and P. R. Kleindorfer, Boston: Kluwer Academic Publishers, pp. 22–41.
- Cremer, H., De Rycke, M. and Grimaud, A. (1995), 'Alternative scenarios for the reform of postal services: optimal pricing and welfare', in *Commercialisation of Postal and Delivery Services*, edited by M. A. Crew and P. R. Kleindorfer, Boston: Kluwer Academic Publishers, pp. 247–267.
- Curien, N. and Dognin, E. (1995), 'Le service universel: quelle valeur? Quel coût? Quel financement?', *Annales des Télécommunications*, 50, pp. 337–347.
- Federal Communications Commission (FCC) (1996), 'Recommended Decision on Universal Service', CC-Docket No 96–45.
- Federal Communications Commission (FCC) (1996), 'The Telecommunications Act', November.
- Gallet, C. and Toledano, J. (1997), 'The cost of the universal service in a competitive environment', in *Managing Change in the Postal and Delivery Industries*, edited by M.A. Crew and P.R. Kleindorfer. Boston: Kluwer Academic Publishers, pp. 304–320.
- Golay, J. and Dobbs, I. M. (1996), 'Costing the universal service obligation: the profitability, distributional and welfare consequences of retaining uniform prices', in *Cost of Universal Service*, edited by U. Stumpf and W. Elsenbast. Bad Honnef, Germany: WIK Proceedings, pp. 41–62.
- Guesnerie, R. and Roberts, K. (1984), 'Effective policy tools and quantity controls', *Econometrica*, 52, pp. 59–86.
- OFTEL (1997), 'Universal Telecommunications Services', Consultative Document (February).
- Phlips, L. (1983), 'The economics of price discrimination', Cambridge: Cambridge University Press.
- Sharkey, W. and Sibley, D. (1993), 'Optimal non-linear pricing with regulatory preferences over consumer types', *Journal of Public Economics*, 50, pp. 197–229.



# PART THREE

## Main policy issues in network industries

### Telecommunications

#### 1. Key technological specificities

The most recent technological evolution has been the sharp decrease in the cost of transmission and switches. The second key development has been the substantial increase in the intelligence of the network, which allows for a richer set of offerings as well as a more efficient utilisation of networks. Furthermore, a large number of potential players appear with railroad infrastructure owners, electricity, gas and water utilities.

Fibre optic cables for long-distance telecommunications have been duplicated in some countries, such as USA, England and Japan, as will probably be the case in continental Europe after 1998. Local telecommunications have also witnessed new developments. Competitive access providers build special access, that is unswitched links, from central business districts and large urban users to specific long-distance companies as well as special access between the long-distance companies' points of presence, and private lines between an end-user's multiple locations. These lines bypass the local network to access long-distance. The local traditional copper wires networks are challenged by alternative technologies. The cable operators have a network in place, but cable TV networks generally have a one-way tree and branch structure that is ill-adapted to the two-way telecommunications services and needs to be upgraded. Cable TV networks have nevertheless started to compete in telephony, for example in the UK by adding a copper twisted pair to their coaxial cable. Wireless services have also developed local area networks (LANs) which are rings connecting computers with outside networks through private branch exchange switches designed for high-speed data applications. Nevertheless the local loop is still perceived as a bottleneck. The link close to customer premises is essentially

a fixed cost and the question remains of the access of all telecommunications actors to this link;

Technological development is changing the location of bottlenecks and a constant attention to the regulatory implications of technical progress is particularly required in telecommunications. Finally, note that, because of network externalities, the multiple networks which proliferate must be interconnected.

#### 2. Policy issues in the transition to full network competition

In the transition phase the entrants will still need to use the incumbent's network to access the final consumers. The major policy issue is then the determination of the modes of access (catalogue of interconnection) and the access prices. The Commission's broad request for access prices oriented towards costs will have to be made more precise. We believe that purely cost based access prices are bound to raise substantial difficulties, creamskimming or price squeezes which will create endless disputes between the incumbent and the entrants. The need for more demand based access prices, may be through the use of global price caps, will probably become clear as competition develops.

Entry into local-call markets may be facilities-based; alternatively, it may develop through resale or unbundling. The pricing of these various alternatives will have to be accurate and in relation with access prices to avoid inefficient duplication of some segments of the network.

In the setting of access charges the main policy issue is how entrants should contribute to the fixed costs of the network they use to originate and terminate their calls. The inclusion of these costs into the access prices, as currently proposed, is unsatisfactory (unless it is viewed as temporary) since it may induce inefficient bypass.

Instead of mark-up on access, a tax on all users of the network (the largest possible base) should be favoured.

To facilitate entry, regulators may be tempted to offer temporary favourable discounts for access prices. This policy is dangerous, as it may induce entry of inefficient operators. It is not clear that regulators have the information required to pick the proper entrants.

The universal service obligations (USO) are a hot political issue in this industry. One issue is how extensive the definition should be — it is a political decision — another is how it should be produced and financed. In the short run, these obligations are often imposed on the incumbents who ask for fair compensation. The desirability of using taxes rather than access charges to finance USO is well understood. The computation of the real cost of USO will remain a difficult problem leading to complex debates through the use of engineering models of networks. The Commission should encourage the development in Europe of such models. An alternative which seems to be favoured by OFTEL is to claim that the benefits of providing USO compensate the costs and to discard the problem.

Two opposite dangers exist in this context. One is that the firms in charge of USO do not fulfil properly those obligations by deteriorating for example the quality of costly social services. The opposite danger is that firms gold plate the servicing of these obligations when they are compensated by cost based schemes. Some form of competition in the delivery of USO is desirable with regulatory attention to quality.

### 3. Long-run policy issues

Assuming that the proper networks have developed at least three main policy issues will remain.

Networks will have to be interconnected if network externalities are to be exploited. The conventional wisdom is that interconnection charges can be left to private negotiations between operators. A couple of points should nevertheless be made.

First, if operators do not agree, regulators will have to step in and some principles for pricing will be needed: cost-based pricing or usage-dependent pricing, symmetric or asymmetric charges, non-discrimination rules, publicity of these agreements.

Second, some attention will be required to avoid collusive agreements which set high access charges to blockade entry for new players and to induce high final prices. Despite the emergence of competition, for a long time to come, it will be competition among a few and competition authorities will be very often solicited. It is easy to predict a tremendous activity in this area within Europe.

Some form of regulation will remain in this industry. The next question is then how regulation should be structured. In particular, provided European countries agree on the definition of USO, maintaining national regulators would in the long run duplicate the complex and rather inefficient regulatory structure of the USA, and generate potential conflicts with a Community level competition policy.

### 4. Current issues facing the Commission

The Commission has produced a draft of a proposed policy statement relative to the application of competition policy to access agreements in the sector of telecommunications (10 December 1996).

First, we have noted in our report that the regulation of access cannot be conceived independently of the status of final prices. Apparently, the Commission's draft presumes deregulation of the final prices. However, it is acknowledged that the development of competition will take time.

The problems arising in the transitional period, when a dominant operator is regulated with price caps on final prices, while at the same time tariffs for interconnection are left to negotiation, should be explicitly addressed.

Second, we find a tension to be discussed between the requirement of independent regulatory agencies and the fact that the Commission can impose on a national government to pay damages to a party whose rights as defined by the European competition policy have not been protected by the national agency.

Third, we note some steps towards demand based pricing that we fully support but which might be at odds with the general cost based approach of the Commission.

Paragraph 92 seems to argue in favour of the ECPR rule as an upper bound for the price of access, if the domi-

nant firm can substantiate its low cost in the market. We have pointed out the limits of this rule as a stand alone rule. However, we think that it may be a reasonable predation test when associated with global price caps.

Paragraph 93 about discrimination seems to open the way for further non cost based pricing of interconnection. Indeed different prices of access may be justified for other reasons than cost namely:

- by the fact that users operate at different ‘levels’ opening apparently the possibility of second order and third order price discrimination;
- by excess capacity and by a better use of infrastructure made possible by other tariffs.

Finally, let us note that paragraph 99 and beyond introduces the notion of a collective dominant position which we find well adapted to the risks of coordination in access prices for which we have developed new theoretical arguments in this study.

A first issue we have raised in the report is what is the appropriate concept we want to measure when we define that cost of USO.

Another main issue is the definition and the financing of universal obligations (USO), (COM(96) 608). A major question which is the object of intense debates in the USA is not discussed: should the cost of service on which to base universal service support be computed from forward looking proxy models or be obtained from embedded costs? A number of difficulties associated with this computation are not dealt with. How to approach stranded assets, how to take with account the uncertainty of demand, how to account for cream-skimming of profitable consumers through mobile phones in high cost areas, how to make compatible the further constraint of uniform tariffs over a geographic area with competition?

Another question is how should the principle of subsidiarity be applied in this context. On the one hand it is not obvious to us that the content of USO should be defined at the European level. On the other hand a number of economic inefficiencies (like cost-based pricing, proportionality principles) seem to follow from the lack of a European regulation.

Finally, let us stress that further research should be devoted to the definition of the appropriate ‘fiscal’ basis for the universal service fund. It is not obvious to us that the basis should be restricted to providers of public networks.

## Postal service

### 1. Technological specificities

The processing of mail typically involves four different activities: collecting, sorting (at several stages), transportation and distribution. The postal network presents a number of specific characteristics which crucially affect the analysis of the various policy issues.

First, the natural monopoly property pertains essentially to one segment of the network, namely distribution. The intuitively appealing property that mail distribution is characterised by increasing returns to scale is confirmed by most empirical studies, even though there does not yet appear to be a widespread consensus on how significant they are <sup>(1)</sup>.

Second, the postal network is ‘people based’; unlike telecommunications, for instance, it does not rely on hard wired assets. Consequently, labour represents a significant share in operators cost (typically around 80 %). Given the status of postal workers, these labour costs are essentially a fixed cost for incumbent operators. In addition, the labour force employed in the postal sector represents a non-negligible share of total employment in many countries.

Third, while demand is expanding in telecommunications, it is expected to at best stable, and probably declining, in the postal sector. Many postal products (like letters, direct mail, etc.) face increased competi-

<sup>(1)</sup> In a study based on French data Cazals, et al. (1997) obtain an elasticity of cost with respect to volume for La Poste in the range of 0.5–0.91 depending on the specification (parametric or non parametric) which is used. This corresponds to a scale elasticity of 1.1 (=1/0.91) to 2(=1/0.5). Christensen et al. (1993) estimate this elasticity of 79 % (USPS data) while Rogerson and Takis obtain 35 % Cohen and Shu (1997) establish the existence of scale economies by comparing the distribution cost of a single operator to the total distribution cost with two operators. Bradley and Colvin (1995) use a test to establish that distribution costs are sub-additive. Finally, in a theoretical study of the distribution technology Panzar (1991) presents arguments which point towards the existence of increasing return to scale in this activity.

tion from electronic substitutes like fax and e-mail. This development is particularly significant because it affects the most lucrative segments of the market (mail originating from corporations) and because the possibility for electronic transmission enhances the possibility of bypassing any regulatory monopoly protection (for instance, bank statements can easily be transmitted to and printed out in a foreign country and mailed from there). Other products, like parcels, are not subject to electronic substitution. However, these market segments are not central in the regulatory debate; parcel-post, for instance, has been traditionally open to competition in many countries. In addition, they rely on production technologies which are, to some extent, separate from those used for letters (and similar products). In particular, the sorting of parcels requires specific facilities and even distribution is often operated on a separate basis.

Fourth, there has been a significant amount of technological progress in mail processing during the last decades (introduction of optical character readers, remote barcode sorting, delivery point sequencing, etc.). A full exploration of these possibilities can be expected to bring about some cost savings. However, the potential for future technological innovations in the sector appears to be limited.

Fifth, and last, there is a strong tradition of uniform pricing (with regard to location and distance) in the postal sector, which is explained in part by transaction costs, but also by political considerations.

## 2. Policy issues

Let us now turn to the main policy issues that currently arise in the postal sector. While the basic problems *per se* are similar to those arising in most other network industries, the appropriate policy response is crucially affected by the specificities of the sector. In particular, the importance of employment in the postal sector (combined with the non expanding demand) implies that any policy which may result in a loss of market shares for the historic operator may be difficult to be accepted on political grounds. Political economy considerations, pertaining to the fact that policy-makers may be reluctant to effect reforms which are likely to be opposed by pressure groups (like unions), though potentially significant, are not the only arguments that support this observation. One can also argue that the avoidance of social unrest constitutes, in itself, a contri-

bution to a public good which ought to be considered in a welfare analysis. In addition, it is quite clear that even on strict welfare economics grounds, the presence of unemployment ought to be reflected in the evaluation of the economic cost of labour employed in the sector. In a context of unemployment, the assessment of the costs and benefits of regulatory reforms thus has to rely on 'shadow wages' which are likely to be below the labour cost that can be assessed on a pure accounting basis.

The relative importance of labour cost has yet another implication for the design of regulatory policies. The fact that the capital requirement is low implies that entry is possible on a small scale and entry barriers for 'cream-skimmers' are low. Consequently, monopoly protection may be particularly hard to enforce and bypass difficult to prevent.

We shall now briefly review the main issues that arise. It should be pointed out that they are all interdependent, so that any classification is somewhat arbitrary.

- One major issue in the debate concerns the nature, the degree and the speed of liberalisation. For the time being, there is already some amount of competition, but it is much less significant than in telecommunications. While there are many advocates for a full liberalisation (after a more or less lengthy transition period), it remains that the natural monopoly in distribution is a potential source of market failure. Consequently, the option of maintaining distribution under monopoly protection certainly ought to be considered. In addition, the natural monopoly argument has to be considered together with the problem of financing the cost of universal service which, as explained below, may also call for some monopoly protection.
- Whatever the degree of competition that will be allowed for by the regulatory environment, it can be expected that the historic operators will retain a dominant position in the market (at least within a foreseeable future) <sup>(1)</sup>. Consequently, there is the issue of whether or not this operator has to provide downstream access to its network and if yes, how 'access charges' ought to be determined. Note that in the postal sector access is usually priced indi-

<sup>(1)</sup> The Swedish and Finnish experience are a good, albeit extreme, illustration of this idea.

rectly, through (less than full) rebates in case of downstream access. The access pricing problem arises in a similar fashion when postal customers (mail order firms, banks, etc.) provide pre-sorted mail (possibly delivered to the destination sorting of distribution office).

Currently the European regulation does not explicitly address this issue. There is some limited amount of downstream access, but it is essentially confined to high volume customers (rather than competing operators) and the corresponding discounts (if any) are usually determined through bilateral negotiation (the terms of which are not typically subject to public disclosure). Overall, the debate appears far from being closed, and it is quite closely related to the other issues we mention. The relevance of downstream access depends in part on the degree of monopoly protection which is granted to some products or some activities. It is also closely related to the issue of universal service to which we now turn.

- Another very significant problem concerns the design and financing of the universal service obligation (USO). The current regulation relies on cross-subsidies combined with monopoly protection for some products (defined on a price/weight criterion). Feasible alternatives include access surcharges (combined with a distribution monopoly) and the creation of a universal service fund towards which all operators contribute.

The first of these arrangements has the advantage that it avoids duplication of the natural monopoly segment of the network; the current system prevents such duplication only for some products (namely those which fall into the reserved sector).

Furthermore, as explained in Part two (Section 4) of our report, confining the cross-subsidies to a reserved sector restricts the 'tax base' in an artificial way and this is likely to bring about a welfare loss <sup>(1)</sup>.

<sup>(1)</sup> In a nutshell, the underlying argument is as follows. It is well known from standard microeconomic theory that the welfare loss (deadweight loss) of a tax increases more than proportionally with its per-unit rate. Consequently, the welfare loss per unit of tax revenue increases as the tax increases. Now, this implies that for a given total tax revenue, the total welfare loss will be smaller if many goods are taxed at a low rate than if few goods are taxed at a high rate. Put differently, the larger the tax base, the smaller the welfare loss. This argument is traditionally presented in the context of commodity taxation. However, it immediately carries over to surcharges levied as part of a system of cross-subsidies which are in essence just a special case of commodity taxes.

Both access surcharges and universal service fund allow to enlarge the set of contributing products (and specifically, to include the products of competing operators) which will tend to lessen the welfare burden that the financing of USO imposes on consumers.

Finally, when the USO extends beyond the reserved area (as in the current system), competition tends to be distorted in the non reserved universal service segments of the market (especially under uniform pricing). This may result in inefficient entry and in the incumbent operator losing market shares, even when its costs are not higher than those of the competitors. The two alternative arrangements, on the other hand, do not create such asymmetries and thus appear to be more appropriate as far as the principle of 'competitive neutrality' is concerned <sup>(2)</sup>.

However, there are also arguments which plead in favour of the current system. The access surcharge scheme may reinforce the threat of bypass and the substitution of monopoly protection by a universal service fund raises credibility issues on behalf of the regulating authorities. Political considerations, including the employment situation also have to be taken into account.

- Economists have long advocated the virtues of incentive regulation (price caps in particular), and most of these arguments also apply to the postal sector. The threat of cream-skimming and bypass and the access pricing problem certainly reinforce the case for incentive regulation, including the possibility of a 'global price cap' (incorporating access charges).

## Electricity

### 1. Technological specificities

Because of the diversity of available technologies, a fairly large consensus seems to exist to say that generation of electricity is potentially competitive, with the possible exception of nuclear electricity. It is also agreed that the high voltage grid on the one hand and the local distribution networks on the other are natural

<sup>(2)</sup> See Cremer, De Rycke and Grimaud (1995), for a detailed discussion of this problem.



monopolies. Generation of electricity may occur within a distribution area or may need the use of the grid to reach the final consumers.

Electricity is essentially non storable and requires centralised dispatching. Also, because electricity can flow in any direction, there are strong gains from the inter-connection of small networks. Then, if a network is financially disintegrated, it still needs an entity of coordination working in the interest of the whole group of firms and users.

## 2. Policy issues

### Competitive behavior in generation

The English pool system through which generators sell electricity has been criticised for allowing generators to charge excessive margins. With a few generators the potential for price manipulations seems high. However, it might also happen that excessive competition make it difficult for generators to recover their fixed costs. This is particularly relevant for the huge investments required by nuclear electricity. The very long-term nature of the required investments (as well as safety considerations in the case of nuclear energy) motivates for some the intervention of governments.

### Grid ownership, nodal pricing, and the coordination of investments

The grid can be government owned, owned by the generators or distributors, or else be owned by an independent entity. The main issue is how to price transmission. The complexity of the externalities created by Kirchhoff's laws raises the question of the possibility in complex electricity systems to price transmission correctly from an economic point of view. The relevant marginal costs are very difficult to compute since the traffic on a given line is highly sensitive to the load on all the interconnected lines.

If such economic pricing had to be given up, the decentralisation of the location of generators should also be questioned. Maybe the solution would be for the grid operator to retain control of locations and let competition of generators operate at specified locations through tenders. Clearly the incentives of the grid operators or regulators should be properly devised.

### Distribution

The regulation of distribution is also complex. Some competitive pressure on distributors may be desirable. This may take the form of allowing some specified classes of customers to buy electricity directly from generators. These bypass possibilities should be made consistent with the obligation imposed on distributors to satisfy demand at regulated prices. Also, the interest of the captive consumers should be protected (the reorganisation of the sector in England has been much more favourable to industrial users than household users).

The attribution of distribution franchises might be performed through auctions. The specificity of the investments requires a delicate balance between competitive pressure and incentives to invest.

### Demand management

Because electricity is not storable, generators are obliged to install and maintain large capacities in excess of demand most of the time. A solution is to flatten the load demand curve by promoting energy savings or energy substitution during peak periods. This requires the installation of specific equipment at the users' location and the implementation of time-of-use depending tariffs. Clearly, particular attention should be devoted to the necessary coordination between generators and customers when the sector is vertically disintegrated. In case of direct procurement contracts, the financing by generators of equipment for energy substitution or saving at the customers' site can be more efficient than the maintenance of excedentary overcapacity, both on private and public grounds. Then it should be encouraged and subsidised.

### Stranded costs

One important issue of the switch from a protected but regulated environment to a competitive organisation is the recovery of the so-called 'stranded costs', that is unamortised costs of prior investments that are scheduled for recovery through regulated monopoly rates but would be hardly recovered under competition. Many electricity utilities are likely to be threatened with bankruptcy if unfettered competition were allowed without a transition period to help utilities in covering stranded costs.



## **Universal service**

Universal service obligations are in general restricted to household users and do not raise difficult issues if they are not allowed to choose their supplier.

## **Natural gas**

We mentioned earlier various problems encountered in the USA and the UK during the transition toward competition in the gas industry: inexperience with a competitive environment, compensation for stranded assets, need for the release of long-term contracts entered into before the liberalisation, and existence of a dominant supplier with substantial market power. We now set aside these (important) considerations and investigate how competition could operate efficiently once these transitory problems are overcome.

### **1. Technological specificities**

It is useful here to recall that the gas industry exhibits many similarities with the electricity industry. The supply side (gas producers, power generators) is potentially competitive. And producers, as in electricity, may find it costly to vary their extraction to accommodate fluctuations in demand. On the demand side lie distribution companies and industrial users, with substantial and sometimes unpredictable daily, weekly and seasonal fluctuations in demand. There is also, as in electricity, substantial heterogeneity in demand, with some users requiring firm service and others being willing to accept interruptability. In the middle lies the transportation system (gas transmission system or pipeline network, power grid). This transportation system is a natural monopoly in two respects. First, its costs are primarily capital costs; pipeline systems' costs vary little with throughput and similarly power grids' short run marginal costs (power losses, reactive power) are only a fraction of their total cost. This usually makes duplication of such systems prohibitively costly.

Second, the efficient allocation of resources and system security call for a centralised operation. This is particularly true for power systems, in which Kirshoff's laws propagate positive and negative externalities of a power inflow or outflow throughout the system. The laws to which gas throughputs are subjected within the pipeline network are somewhat simpler, as they result directly from the matrix of inputs and offtakes at the various

entries/exits of the system and from the capacities of the branches of the pipeline network. But it is generally accepted that a centralised operation or at least a very tight coordination of the pipeline network's element is desirable. The treatment of the transportation bottleneck is the key to the opening of competition in both industries.

There are however two differences between electricity and gas <sup>(1)</sup>: electricity is not storable while gas is. In this respect, it is important to distinguish between (at least) two types of storage facilities: seasonal storage facilities on the one hand, and diurnal and emergency back-up storage facilities on the other. Seasonal storage facilities (such as Rough, a depleted gas field in the North Sea) allow suppliers to produce gas more effectively ('at a higher swing factor') and are not meant to offer peak gas supply. They are basically a cost-reducing or efficiency-enhancing input into the production process. Although there is potential competition in the provision of storage, geographical or historical conditions may make existing storage an essential facility to which suppliers must have access in order to compete effectively. In contrast, storage facilities intended to smooth daily peaks and to meet emergency requirements (low and high pressure gas holders, etc.) are an integral part of the transportation system in the same way reactive power belongs to the power grid. In other words, such storage substitutes for pipeline and compressor capacity and helps balance the system, and thus cannot be easily unbundled from transportation services.

In contrast with the electricity grid operator, a pipeline operator need not always call on suppliers to provide additional input or to restrict their supplies, or interrupt customers, but it can also maintain system balance by using such storage facilities. This of course does not mean that some suppliers should be denied implicit access to storage facilities by being prevented from offering firm service to their customers (that is, by being denied firm access to the transportation system), but rather that diurnal and emergency back-up storage should a priori not be purchased separately from transportation.

<sup>(1)</sup> Another difference between gas and electricity is that gas faces competition from alternative energy sources on basically all market segments.

Another difference between gas and electricity is obviously the very limited number of gas producers (Russia, Norway, Algeria, and the Netherlands, with the UK exporting very little). The efficiency gains of more competition in production will therefore be fairly limited, as no newcomers will be able to enter in practice.

## 2. Policy issues

We certainly view the introduction of competition in the gas industry as a very positive development. We feel however that the reforms proposals are often guided more by a voluntarist competitive perspective than by some well-accepted framework for efficient competition. Indeed, and like in the electricity industry, there has been a dearth of in-depth reflections in academia and policy circles as to what an efficient organisation of the industry should entail. Such industries are not drawn at random, but have specificities that suggest the need for thorough investigations of how competition is to be introduced. Within the limited scope of this report and calling wholeheartedly for careful analyses, we content ourselves with a list of issues that have to be addressed in the process of designing an efficient industry organisation.

1. Vertical structure. As in the electricity and other industries, a recurring question is whether the bottleneck owner should face line-of-business restrictions, that is be prevented from entering potentially competitive segments. In the historically vertically integrated electricity and gas industries, this question can be phrased in terms of the desirability of the divestiture of the incumbent's supply and transportation units (and for gas, of seasonal storage facilities). The rationale for structural separation is that it may be easier to establish a level playing field among firms in the competitive segment if none of them is integrated forward into transportation. The potential cost of divestiture is a lack of coordination of supply and transportation investments. It is however important to note that the desirability of divestiture depends on how the system is designed to operate under vertical separation and vertical integration. This brings us to the next three items on the list.
2. Short-term allocation. Taking investments in production and transportation as fixed in the short term, we must look for mechanisms that select the lowest cost suppliers and provide customers with

the right price signals. In a vertically separated industry, the short-term allocation may be obtained through a system of, say, daily bids by customers and suppliers into the pool/transportation network. There are various ways of organising such 'double auctions' and we will not get into details here. Let us just point out a couple of relevant considerations. First, as in the electricity industry, prices should optimally differ across nodes so as to reflect transportation constraints and provide customers and suppliers with the right signals. This implies for example that the concept of notional path linking the transportation charge to the shortest distance between the input and the offtake point is rather meaningless; and that charges should be lower (if positive at all!) when the flow goes counter the prevailing system flow and relaxes the transportation constraints (in the gas industry, this is called 'backhauling'). While the desirability of nodal pricing is well-known, we of course acknowledge that the computation of nodal prices is not trivial; but we feel that their use is important for economic efficiency.

In the case of a vertically integrated transportation system, efficient access prices must be designed so as to encourage efficient entry while inducing the vertically integrated operator to invest and maintain the network properly and not to try to deny rivals' access to the bottleneck. We refer to our earlier and comprehensive discussion of the matter.

3. Long-term coordination of investments. In a non-competitive, vertically integrated system, the overall level of investment in transportation and production and its locations was presumably well coordinated within the monopoly firm. The introduction of competition raises the important question of whether investment in capacities, which represents a substantial chunk of industry cost, will not exhibit serious rent seeking and waste. There is a substantial and largely unsettled debate in the electricity industry in this respect (which involves various suggestions as to the booking of transportation facilities, property rights on transmission segments, incentive mechanisms for a monopoly transportation owner, obligations to build additional capacity, use of insurance mechanisms such as contracts for differences, and so forth). A similar (and perhaps less developed)

debate is not settled in the gas industry either. Yet, we feel that the issue of the coordination of investments must be thoroughly addressed by the competitive reform process.

4. Incentive mechanism for the transportation owner. Relatedly, the transportation owner must be given proper incentives to maintain and develop the pipeline network. A theoretically desirable regulatory solution is to impose a price cap on a basket of nodal prices (to then be computed by an independent system operator). Such a pure price cap however has some drawbacks when it comes to regulating the network owner's rate of return. As usual, mistakes in the computation of the overall cap may result in undesirably large rents or conversely in the absence of cost recovery. Also, and a particularly relevant point in view of the fact that much network costs are capital costs, the network operator may be concerned about a potential regulatory expropriation of its investment through sharp reductions in the overall cap once investments are sunk. Cost sharing (with suppliers and customers) — or in the extreme some guaranteed 'fair rate of return' — fare better in this respect although they reduce the transportation owner's incentive to operate efficiently.

## Water

### 1. Key technological and economic specificities

#### 1.1. Local sub-networks and scale economies

Resources in water differ from one country to another. Furthermore, these resources may be unevenly spread within each national territory.

Given that transport costs are relatively high, one generally observes the existence of a large number of local networks (cities, metropolitan areas, counties, etc.) which, often, are not interconnected. However, interconnection may prevail when it is necessary to provide water to regions with a particular need. This is, for example, the case of Flanders in Belgium and some Spanish provinces.

As water is a renewable resource, each local network is in fact composed of two sub-networks:

- a distribution network which provides water to users;
- a water purification network: a sewerage system that collects used water, purifies it, then recycles it through the network or rejects it in nature.

In each of these sub-networks, water goes through pipes. Hence, the activity of water supply is inherently characterised by economies of scale (according to the 'two-thirds rule'). Natural monopoly seems then to be the fundamental economic structure of this activity, and indeed, we observe that, in all countries, each of these sub-networks is managed by a single firm. There does not seem to exist scope economies between these two sub-networks.

#### 1.2. Global management of the water resource

While, in some countries, such as those of the Mediterranean, quantity of water is an issue, quality is an important matter in all countries. Indeed, the externalities caused by the pollution of water raise a great problem of management. One classifies these externalities according to the polluting agent.

Households' used water generally seems to be well recycled. However, a potential problem might arise when rain drains chemicals, due to automobile traffic, towards rivers. In the case of industrial firms, the pollution is, in most cases, observable and hence regulated through standard methods (taxation, subsidies, norms, etc.).

In the case of agriculture, the pollution is essentially due to the use of fertilisers and pesticides. Here, the pollution is not punctual as in the case of industry, but rather diffuse. In particular, it may affect the underground stocks of water. Furthermore, measuring the degree of pollution by individual polluters is generally prohibitively costly and imprecise, which makes regulation of this pollution a difficult task.

Two further points deserve attention. First, depolluting units produce a by-product that may pose a large problem of processing and storage. Second, the management authorities might face some peak load phenomena. Those can arise in periods of irrigation during the summer, during particularly stormy weather episodes, and when the level of demand of water is particularly high due to holidays.

## 2. Policy issues

### 2.1. Regulation of local networks

While water pumping, billing and revenue collection might be organised in a competitive way, distribution and sewerage are natural monopolies that necessitate the choice of the firm(s) and the regulatory scheme under which it (they) will operate. A possible way of selecting the operator is through auctions.

Following the choice of the operator, several standard problems of regulation have to be dealt with, most notably the trade-off between the need to give the firm incentives for efficiency and the desire of the regulator to extract the monopoly rent. A standard illustration of this trade-off is given by the difference between the price cap and the cost plus contracts.

However, a specific problem arises here because the network infrastructures are underground. The evaluation of the precise quality of the infrastructures, in particular when they are old, is quite difficult. Because of this great uncertainty, privatisation (which requires an assessment of the value of the infrastructures) is sometimes difficult to implement. Moreover, if the regulator chooses to put in place a concession contract, one must assess *ex ante* the extent of the investments needed to replace the used elements of the network and to maintain them. Finally, the monitoring of these necessary projects is a difficult task as well. This lack of *ex ante* information can likely explain the length and the complexity of the concession contracts that are observed in this sector <sup>(1)</sup>. Furthermore, the contract must incorporate some rules of adjustment as new information becomes available.

### 2.2. Global management of the resource

As discussed above, there is a substantial externality problem due to pollution and one might attempt to use standard tools to alleviate it. However, in some cases, one does not observe the level of pollution of individual agents but only that of a group of agents. In those cases, the pollution charges can only be made dependent upon imperfect proxies such as the levels of individual water consumption. Moreover, as water runs along rivers that often cross several regions and/or countries, the issue of

the appropriate geographic region under jurisdiction is an important question. A certain degree of coordination among the regulatory authorities of the various regions and/or countries is necessary. In this respect, various schemes have been adopted by European countries. For instance, the UK has a centralised system, whereas the German system is more decentralised. In France, the SDAGE ('Schémas Directeurs d'Aménagement et de Gestion des Eaux') may be viewed as an attempt to take care of this coordination problem.

## Urban transport

### 1. Characteristics

#### 1.1 Technology

Most econometric studies show that urban transport networks exhibit large increasing returns to scale. Starting from this result, care ought to be taken to understand how these economies of scale are obtained. In urban transportation economics, one distinguishes between economies of density that are produced by expanding the density of output, and economies of size that result from an increase of the spatial scale of output. To operate this distinction, it is usual to distinguish between measures of final outputs, such as passenger-kilometres, that account for user time, and measures of intermediate outputs, such as seat-kilometres, that provide the potential production. Empirical studies suggest that intermediate outputs are produced under constant returns on average. Indeed, if one is able to exhibit increasing returns for small firms, they seem to be exhausted for a bus operating firm covering a large city. In addition, as capital stock (fleet size or trackage for rail networks) is quasi-fixed in the short run, one may observe returns to scale which are probably much higher than in the long run. Now, as the capacity utilisation of vehicles can be increased as demand grows (people can stand and crowd to some extent) and as the waiting and transfer time spent by the passengers can be changed drastically, there are considerable returns in producing final outputs such as passenger trips.

The crucial technological characteristic is here the time provided by passengers to the public transportation modes which have to match timely and spatially desired trips to vehicles, trains or buses. Two conclusions can be drawn from this fact. First, the central issue in urban transport is congestion, in which users impose costs to

<sup>(1)</sup> An efficient way to deal with this uncertainty might be to auction off the concession.

others by increasing their waiting and transfer times. The questions of interoperability, intermodality and interconnection are just other ways to deal with this issue. Second, the industry must be arranged in order to capture the economies of scale at the level of the final output, i.e. the achieved number of trips. One may well create a vertically integrated firm to which users are committed through the network (the physical layout of bus stops, the ticketing service, for instance); then this firm would be a natural monopoly. Or one could operate the urban transport system around several firms as long as they provide adequate combined services, i.e. they manage externalities among themselves; then the increasing returns are kept at the industry level.

## **1.2. Regulation**

The dominant organisational form for providing urban transportation services to individual users in developed countries is public ownership. When provided by private firms, the services are heavily regulated. Regulation bears on the network structure, the capacity and quality of service, the fare structure, the cross-subsidies, the level of investment, the financial structure, etc. The common feature of the urban transport industry is that it is heavily subsidised, with transfers from the authority ranging usually from 25 to 75 % of operating costs.

The efficiency of these regulatory arrangements is regularly questioned. While empirical evidence supports the view that private operators are more efficient, these results are hampered by various effects. For instance, it is argued that publicly owned firms must support undue costs or are forced to offer services that are clearly unprofitable. As a matter of fact, cost efficiency is presumably more related to management incentive systems than to the form of ownership.

There are three cases of interest with respect to the regulatory issues. First, the long experience of contracting between local political authorities and private operators in many French cities shows that these contracts have not solved the financial problem of urban transportation systems. Recent studies indicate that the type of contract matters, with fixed-price-type contracts being associated with lower operating costs. Second, evidence from deregulation and privatisation of urban transport can be drawn from the UK after the British Transport Act of 1985. Large cost savings have been achieved and subsidies have been strongly reduced. On the other hand, fares have risen while passenger-trips and rev-

enues have decreased. The experiment seems to have been more successful in London than in the other metropolitan areas or in the shires, and for buses than for urban trains. Third, full deregulation of urban transport and in particular, urban railways in Japan appears to be a success for most of the analysts.

Other experiments can be quoted to show that the regulatory environment is quite changing in urban transportation systems of developed countries. The key issue here is the diversity of authorities who could intervene in the organisation and provision of public transportation at an urban level. In most countries, there are several layers of regulators. For instance, the central government plays a role by fixing a cap on the change in average prices, while the local authority (at a regional or city level) decides on other regulatory issues. This political organisation does not seem efficient always.

## **2. Policy issues**

### **2.1. Liberalisation and regulation environment**

There is an increasing support for private participation in providing urban transit services. However, choosing the adequate scheme of deregulation and regulation is still debated. The two polar cases are the use of contracts by a more or less politically independent agency in charge of the overall organisational coordination for the service, or the competition among private and public firms, with no restrictions on entry. When restriction of access exists, competitive tendering seems to be an efficient way to allocate the rights of access, with or without an agency in charge of monitoring the services offered, and with or without subsidy.

Three remarks shed light on this debate. First, a system of urban transit organised around the competition of several firms is hardly conceivable. Indeed, this raises the question of using the same set of transport nodes and operating different types of vehicles and services. Building a system able to benefit from the overall increasing returns of final outputs is not so easy. For instance, how should one organise in a competitive way the private provision of urban transport, ticketing and scheduling, which are sources of economies of scale and scope? Second, urban transport is not a homogeneous good, with some services being more profitable than others. Express commuting services between rich and/or populated suburbs and business centres are presumably a target for new entrants. This problem is



related to the issue of ‘cream-skimming’ when obligations of universal services are imposed on some providers. Third, urban transport markets do not seem to be contestable. Then one cannot rely on potential competition to regulate the monopoly service and incentive schemes are needed.

Is there a scope for a Community guideline on this issue? As already noticed, in many European countries, urban transport is regulated through different levels of political entities which could create potential sources of conflict, hampering the adoption of a more efficient regulatory system. Here research plays an essential role. Favouring the developments and the implementation of the latest scientific, technological and organisational views will help enhancing the efficiency of urban transportation systems. While research programmes initiated by DG VII are motivated by a similar concern, urban transport may benefit from the analysis of policies adopted in other sectors, like telecommunications. Indeed, intermodal competition is very much like competition in access technologies on a whole telecommunications network.

## **2.2. Environmental issues, in urban transportation technologies, and intermodal competition**

Whatever the regulatory scheme, regulators will face the need of improving intermodal passenger transport within urban centres in a door-to-door chain. Upon the availability of technologies fitted for implementing intermodality, this task is not so easy to accomplish from a political and economic point of view. Developing public urban transport is hardly reconcilable with the overwhelming use of private cars. However, as the environmental impact and the congestion costs become a central issue, upper level decisions are needed. The fact that urban transport is largely subsidised through taxes and that the pricing structure does not usually reflect marginal costs of providing the services (i.e. prices are uniform, independent of time and distance), complicates the situation. Presumably, local authorities would face the trade-off between improving the market share of public urban transport while saving space and reducing pollution at the price of higher subsidies, and imposing drastic conditions on the use of private vehicles. Of course, technological progress may help solving the problem and should be encouraged. In the meantime, Community guidelines are probably needed on these matters.

## **Air transport**

### **1. Technology and networks**

There are two different markets in the air transportation industry: passenger transport and freight. Here, we will only consider passenger transport, which forms the greatest part of the income of the industry. This is by far the most important part of the industry, and a large proportion of the freight is carried in the same aeroplanes as passengers.

The air transportation industry has three major components that require substantially different economic policies:

- airlines;
- traffic control;
- airports.

#### **1.1. Airlines**

In the standard, elementary, technical sense, there does not seem to be many increasing returns to scale in the airline industry. Flying two planes from one airport to another does cost not very much less than twice the cost of flying one plane between these two airports. However, the marginal cost of a passenger on a flight which is not full is very close to zero, and hence the cost function that links the number of passengers flown to the cost has jumps. This explains the importance of the drainage of consumers to long distance flights, and the development of feeder routes in recent years. On very small routes, the situation is somewhat different, as the fixed costs of a minimum crew and ticketing facility will already represent a sizeable portion of the total expected revenues, and increasing returns to scale become important.

Other facts indicate that there are benefits to size within the industry.

- Consumers are sensitive to the frequency with which an airline flies between two airports, in great part because they cannot always predict the time at which they arrive at the airport, hence the rapid recent development of shuttle type service (see for instance the new SAS Express service).



- Connections are easier between flights of the same airline, and passengers prefer travel with the same airline from one end of a trip to the other (it is in order to give them the feeling that they are doing so that codesharing has been introduced). An airline with many routes has therefore an important competitive advantage.

Although market power often stems from increasing returns to scale, in many industry it is also caused and sustained by product differentiation. This also holds true for airlines, who try to differentiate their products by offering different and better service.

Finally, difficult issues arise because the most important strategic resources of airlines, time slots at the busiest airports, are public property <sup>(1)</sup>, and the administrative allocation of the rights to use them creates distortions.

### **1.2. Airports**

A large array of services are provided at airports: the most visible is, of course, the management of air traffic when landing and taking off, but there is also maintenance of aircraft, baggage handling, ticketing, restauration, parking, car rentals, and so on. The landing and taking off part forms a natural local monopoly as the fixed costs of building the runways and the associated infrastructure are very large. Only in the largest towns is it economically advisable to build more than one airport. The fact that the activities of airports are very diversified, and that the degree of increasing returns to scale is not the same in all the activities is shared by the nodes of very few network activities (a prominent example would of course be train stations).

Airports also share with train stations the fact that they are intermodal nodes, at the intersection of several networks, the airline network and the urban transportation network.

### **1.3. Air-traffic control**

Air-traffic control is in charge of coordinating the flight paths of the planes, and this is by essence an industry in which there is a natural local monopoly, in the same way that there is a natural monopoly for provision of

traffic lights at any given intersection: no two systems can control the same airspace. At the present time there is a national monopoly for this activity. Given the importance of international flights, proper procedures need to be put in place to ensure the coordination between these national monopolies.

## **2. Policy issues**

### **2.1. Airlines**

Historically, the airline industry has been the first network industry to be deregulated, in the United States in the 1970s. There seems to be no major reason to reverse this trend, but even if this conclusion is accepted, there do exist important policy issues that need to be addressed within that framework.

The first policy problem is the allocation of slots at airports, and maybe in some parts of Europe the allocation of air routes. The liberalisation of the industry in Europe can only bear fruits if entry on the main routes is not impeded by the difficulty in obtaining landing slots. As discussed in Part one, the use of auction type mechanisms for the allocation of the right to use slots should be explored, along with consistent regulations of the usage rights of slots. The extreme importance of this issue for the industry makes it a very difficult policy problem, both because mistakes in the allocation of slots will be very costly and because political pressure is very strong.

The second policy problem is the necessity for a strong competition policy in that sector. Some economists believed that competitive pressure would be sufficient to prevent airlines from exploiting their monopoly positions. The data made available by the US experience of deregulation shows very clearly that this is not the case. For instance, the prices of flights from US hub airports dominated by one airline are higher than the prices of similar flights from airports where several airlines compete. More direct tests have shown that, contrary to what the contestable market theory would predict, potential competition is not as potent as actual competition in bringing prices down. Among others, frequent flyer programmes, code sharing, and the management of computer reservation systems should be examined as potential barriers to entry and/or tools for collusion.

Although increased competition is probably desirable, it does create problems of its own. The airline industry has a number of features, differentiation of products,

<sup>(1)</sup> This is not entirely clear. Slots are protected by grandfathering clauses, the trade press takes it for granted that there is a 'grey market' in slots. In some respects installed airlines use slots as their private property.

need for coordination between competitors, the purchase by consumers of complementary products from competitors, which imply that even with competition the allocation of resources will not be optimal. For instance, the economic literature has shown that airlines may have a tendency to provide more flights at peak hours than is socially optimal, and not enough at other times. Some of these distortions can be counteracted by appropriate policy intervention (for instance, subsidies for off-peak flights), but others will remain. On balance, we do not believe that the presence of these distortions is argument enough to go back to regulation, as long as competition policy is attentive. Indeed, these distortions do not seem to be more important in the airline industry than in other industries where regulation is not deemed necessary.

Finally, as seen above, for routes with little traffic, the returns to scale cannot be neglected. Under these circumstances it is well known that private firms may find it unprofitable to provide service, even in circumstances where it would be socially optimal to provide it. It can be desirable to subsidise service, to authorised by article 4 of Council Regulation (EEC) 2408/92 and is done in France through the 'fonds de péréquation'. The contract with the airlines receiving this subsidy should specify frequencies and should probably put an upper bound on the level of at least some fares. They will have incentives to provide good service by the desire to maximise their revenues. There is a specific investment which is necessary in order to open a route, and therefore the contract should be for a sufficiently long period. The airlines receiving these subsidies should be selected on the basis of an open and non-discriminatory auction. Ideally, the subsidy should be levied through the general budget, but a general tax on air travel is probably not too bad a second best.

## 2.2. Air-traffic control

As discussed above, air traffic control is a natural monopoly. Many governments have recently put independent agencies in charge of this activity, some are even trying to privatise it. This should not be interpreted as the introduction of competition, which cannot exist, and the policy discussion should bear on the best method of providing this special type of public service. Depending on national traditions and legal frameworks, different solutions might be better adapted to different national environments. Given the fact that there is a very large amount of specific capital, both physical and human, concessions may be difficult to implement.

The Commission has suggested the need for regulation at supra-national level (Communication of 11.9.1996 on services of general interest in Europe, section 65). It is beyond the scope of this report to compare the existing national systems with a hypothetical European system.

## 2.3. Airports

Although airports are natural monopolies in some cities, in a number of them there is some degree of competition (London, Paris). Some thought should be given to the possibility of increasing competition. For instance, it could be beneficial to put airports serving the same city in competition with each other rather than having them managed by the same entity. It should be noted that although airports have a monopoly, airlines that use it have a strong incentive to push them towards greater efficiency, as this affects the demand for their services, and they should be given a say in their management

Despite this natural monopoly component, there are ways in which competition can be introduced in auxiliary services, thus decreasing the possibility for the airport to extract monopoly rents. This point is already taken into account in European policy. Care should be exercised to ensure that competition is effective. Indeed, the nature of the services which require much coordination between rivals (as for instance in luggage in transit) may facilitate collusion, and it may be difficult to distinguish between technically beneficial coordination and anti-competitive activities.

It goes without saying that collusion between airports and one airline should be prevented, as much as possible.

# Railway transport

## 1. Specificities

Traditionally, railways have been viewed as a natural monopoly, hence, vertically integrated monopolies are widely observed in the sector. In fact, only the networks and the station activities have the natural monopoly. Competition could be introduced in the production of the services of transport, but actually, the main sources of competition are the other modes of transport.

From a technological point of view, it is the advances in the other modes of transport (planes, cars, trucks) rather

than innovation in the railways sector itself which explain many characteristics of its evolution. High speed trains have, indeed, constituted a major technological advance, but it was not sufficient to stop the decline of the rail. Furthermore, this innovation has concerned only the passenger traffic, while freight might be, in Europe, the only future of the rail. For the freight transport, the technological innovations explain the development of the railways in the USA as much as deregulation of the sector. The development of the container technology is the major technological change for the freight transport by train.

From a technical point of view, no European network per se exists in the railways sector: the problems of rail tracks, gauge for tunnels bridges and the problems of electricity supply make the traffic of trains through the European continent very difficult as it requires heavy operations at the borders. This is the general problem of the interoperability of the rolling stock. This problem is particularly severe in railways because of the lack of flexibility and the difficulty to obtain marginal improvements. While European networks exist, at the present time, no European network is available. This explains much of the high cost (financial or in terms of time) of some transport services by train and the significant decrease in demand for some passengers services and freight transport. However, Article 154 of the EC Treaty, Directive 96/48/EC on the trans-European high-speed rail system, and more generally Decision No 1692/96/EC on Community guidelines for development of the trans-European transport network constitute an appropriate legal framework. But their application to railways is not effective.

From an economic point of view, the reduction of rail's share in passengers and freight transport markets is spectacular in Europe. During the last 25 years, the transport of passengers by train (measured in passenger-kilometres) has grown by 25 %, while transport by car has increased by a factor of 2.2. Freight transport by rail fell by 30 % and transport by lorry was 2.5 times larger (in tonne-kilometres). The predictions are extremely pessimistic: in the next 10 years, the share of the railways could be as low as 4 % for the transport of passengers and 9 % for freight. If these forecasts prove correct, the very survival of this mode of transport could come into question. Indeed, this question would effectively arise if the economic organisation of this sector remains unchanged.

However, as a mode of transport, the rail allows to solve many environmental and congestion problems related to the sector. This could particularly be the case for the passengers traffic with corridors between large urban concentrations. The regional passenger traffic by trains is also a good solution to the congestion and pollution urban problems generated by individual transport modes. Railways are also more efficient than the road for the freight transport when supply is adapted to demand by the shippers. In this case, railways could solve many of the questions raised by the traffic of trucks: safety, congestion, etc. But it is clear that the present state of the sector prevents such a positive evolution for the railways.

In any case, the situation is not uniformly catastrophic. In the last 10 years, the transport of passengers has increased by 55.2 % in the Netherlands, 32 % in Italy and Austria, while in Portugal the passenger traffic decreased by 16 % and by 10 % in France, in spite of the introduction high speed trains. For freight, the situation is even more heterogeneous, since one observes strong rises in Portugal (+ 68.9 %), Italy (+ 32 %) and Austria (+ 21.5 %), and very strong reductions in Germany (– 42.3 %) and France (– 17.9 %). Meanwhile, railways were strongly developing for passengers in Japan and for freight in the USA where this mode is currently dominant.

The organisation of the sector is very heterogeneous. In most cases the sector is characterised by a State-owned monopoly and often by important social problems. The sector is entirely privatised only in Great Britain (except for the tracks) and, in some cases, it is possible to observe the emergence of small private companies in the regional part of the sector, like in Sweden. In other cases, like Germany and Italy, the organisation of the companies, in principle, should allow privatisation, with a complete separation between the different activities in the sector: network, passengers, freight, etc. In many cases, passengers transport and freight transport are separated from network activities, and could support some degree of competition with new (private or not) operators.

Of course, the network is everywhere a natural monopoly. But it would be false to conclude that the decline is faster where the companies are State-owned or where the different activities are not separated, excluding the introduction of competition. For example, the highest performance is achieved in Sweden, in terms of productivity and revenue by employee, where the State is

strongly involved, but also some degree of competition between operators on the network exists. The French national company SNCF, which is an entirely State-owned monopoly (for all the activities of the sector) is, respectively, at the fourth and sixth position in Europe according to these criteria of performance. In fact, many national particularities, not only in the railways sector, but also in the geography and in the characteristics of the other modes should be considered in order to compare the performances of the different companies. It is not easy to construct appropriate comparisons between alternative organisations of the railways sectors in Europe or elsewhere in the world. This remark holds for the comparisons with Japan and USA as well.

Except for some small networks in Europe and also in the USA and Japan, the burden of the debt is heavy everywhere. The debt could be a problem if the networks were opened to competition, because this competition would be distorted against the 'old' companies which have in many cases supported entirely the infrastructure investments. At the present time, no competition at all exists in the railroads except in the UK where there is *ex ante* competition through franchising. When the application of Directive 91/440 EEC will be effective in the Union, the context will be very different, first because the debt will be transferred to the States, and secondly because some degree of competition will be in principle introduced on the networks. But Directive 91/440 EEC will produce very few effects in terms of competition as long as the two complementary directives (95/18/EC and 95/19/EC) are not effectively applied. Actually, no competition at all exists on the rails. So the level of the debt should not be considered as alarming in itself because it is often used to finance investments which should be sufficiently productive to allow its repayment. In fact, the debt is a problem only when the costs of a company are so high, and its productivity so low that it is impossible to repay it and recover the current costs without the assistance of the public authorities. And when this assistance is quasi-automatic (which is most often the case in the railways sector), it is not sure that the investment choices are correct, because this assistance may induce a strong tendency to underestimate costs and to overestimate the traffic in the evaluation of the investment projects.

## 2. Policy issues

All the policy issues raised by the railways industry revolve around the question of how to introduce a cer-

tain degree of competition in the sector in order to improve the efficiency of the companies and to adapt supply to demand, not only in quantity but also (and importantly) in quality. Technical problems are often invoked in this sector in order to maintain the monopolies or to avoid the introduction of some degree of competition. In fact, the introduction of the market forces in this sector seems necessary for the future of the rail. Furthermore, the issue becomes even more crucial when one observes that the rail transport is one of the best responses to congestion and pollution problems.

The separation of the different activities is realised or will be rapidly realised in Europe, with the application of Directive 91/440/EEC. It is a necessary condition for introducing competition among potential operators on the network. The principal questions here are: which authority will be in charge of the regulation, and how are the access pricing to the network and the allocation of the slots on the different portions of the network to be determined? Actually, access pricing, when available, is entirely cost-based and this will be the source of systematic disputes between operators if competition becomes effective. It is clear that a demand-based access pricing will be necessary to obtain some improvement in the structure of the supply of rail services. Then pricing will be a means to determine priority on the rails and to supply more adequate services. Of course, some obligations like universal service must be taken into consideration to establish the priorities. Clearly if the network activities are organised by an independent authority, distinct from the operators of the network, it will be easier to organise competition fairly. It cannot be the case if the same company manages the network, determines access pricing, allocates the track capacity and supplies itself transport services on this network. In this case, the incumbent will have the possibility to indirectly prevent the entry of new operators, or to prevent an efficient supply of new services, notably if the incumbent is not competitive in the production of these services. A related problem is the problem of interoperability between the different national networks. At first sight, one would think that it is only a technical problem, but in fact it could be a means to deter entry into each other's markets. It is crucial to organise this interoperability and to construct common economic criteria allowing to appreciate objectively when interoperability can be implemented with 'reasonable' costs, and when it is impossible.

Concerning freight transport services, improvements could come from dedicating parts of networks to freight services only. If this is not feasible, the authority in charge of the network should give the priority to the freight on some links. Indeed on some links, the mix of traffic (freight, passenger on short and long distance) could cause congestion problems. In this case, passenger traffic usually has priority. Hence it is not possible to obtain high quality service for the freight transport by rails on such links, and other modes of transport (road essentially) are more competitive. The viability of the trans-European freeways is strongly related to this question. The concept of 'trans-European rail freeways' seems appropriate, but it is far from being accepted everywhere. The ability to transport freight (and eventually trucks) through Europe on dedicated or quasi-dedicated links could enhance supply of freight transport by rail. The allocation of the slots on these 'freeways' will be of course strategic. The intermodality is a strongly related problem. The most interesting solution using intermodality is certainly the container technology since the container can be transported indifferently by ships, trucks and trains. But, in order to have an efficient use of this technology, a sort of 'one-stop-shop' must be organised to provide charges and services for operation on the whole corridor. Because this technology can potentially use all the modes (ships, trucks, trains, and eventually planes), cooperation between companies in the different transportation sectors is required in order to supply such a 'one-stop-shop' service to the shippers.

The substitutes to the container technology do not offer the same guarantees for competition because they implicitly give an advantage to road transport: the same trucks are charged on the train and transport the freight until the final destination, and it is clear that the access

pricing to the rail network should be very low in order to make the cost of this transport system competitive. Except on specific corridors (like in Switzerland or Austria for example), where congestion and pollution problems are very important, the container technology seems to offer better characteristics for a competitive transport service. In any case, the transfer of a significant part of the European freight transport to the rails should bring an improvement in the cost of this service for the customer as well as a solution to the congestion and environmental problem. The main objective should be the improvement of the welfare and not the rescue of the railways in a strict sense. The challenge is though, to get the market forces and technological innovation to play a major role in this rescue. Hence, the policy measures must allow for an efficient introduction of market forces and new technologies in this industry.

Another important policy issue in the railways sector is the social impact of the change in the organisation of the sector. In many cases the traditional situation is a State-owned monopoly with many social advantages for the employees. Of course, these advantages may make these companies non-competitive if competition occurs. In fact, this can prevent the introduction of some degree of competition on the rails if the States want to avoid social movements. An interesting solution was adopted in Germany by DBAG. A new structure was created, and the employees of the old railways companies had to choose between being directly employed by DBAG under private sector labour conditions and joining the new structure with their civil servant status. DBAG can use the services of these employees under the private contractual conditions, the federal government subsidising the difference. With this mechanism, DBAG does not support non competitive cost conditions, and its former employees preserve their social advantages.



# References

- Bradley and Colvin (1995), 'An Econometric Model of Postal Delivery', in *Commercialization of Postal and Delivery Services: National and International Perspectives*, edited by M.A. Crew and P.R. Kleindorfer. Kluwer Academic Press, pp. 137–153.
- Cazals, De Rycke, Florens and Rouzard (1997), 'Scale Economies and Natural Monopoly in the Postal Delivery: Comparison between Parametric and Non Parametric Specifications', in *Managing Change in the Postal and Delivery Industries*, edited by M.A. Crew and P.R. Kleindorfer. Kluwer Academic Press, pp. 65–80.
- Christensen, Christensen, Guy and O'Hara (1993), 'USPS Productivity: Measurement and Performance', in *Regulation and the Nature of Postal and Delivery Services*, edited by M.A. Crew and P.R. Kleindorfer. Kluwer Academic Press, pp. 237–255.
- Cohen and Shu (1997), 'A Measure of Scale Economies for Postal Systems', in *Managing Change in the Postal and Delivery Industries*, edited by M.A. Crew and P.R. Kleindorfer. Kluwer Academic Press, pp. 115–132.
- Cremer, H., M. De Rycke and A. Grimaud (1995), 'Alternative scenarios for the reform of postal services: optimal pricing and welfare', in *Commercialization of Postal and Delivery Services*, edited by M. A. Crew and P. R. Kleindorfer, Boston: Kluwer Academic Publishers, pp. 247–267.
- Panzar (1991), 'Is Postal Service a Natural Monopoly?', in *Competition and Innovation in Postal Services*, edited by M.A. Crew and P.R. Kleindorfer. Kluwer Academic Press, pp. 219–228.
- Rogerson and Takis (1993), 'Economies of Scale and Scope and Competition in Postal Services', in *Regulation and the Nature of Postal and Delivery Services*, edited by M.A. Crew and P.R. Kleindorfer. Kluwer Academic Press, pp. 109–127.