1 Introduction

1.1 A new and fast-growing industry

A series of features makes the mobile telecommunications industry an interesting field of investigation for economists: the industry is experiencing very fast market growth combined with rapid technological change; regulatory design in setting market structure is playing a very important role; and oligopolistic competition is unfolding under various forms. The number of subscribers to mobile networks is growing at a rapid rate on a worldwide basis, as shown in figure 1.1. During the 1990s the number of mobile subscribers worldwide increased by an annual rate of 50 per cent. An important year was 2002, when the number of world mobile subscribers for the first time exceeded the number of fixed lines. The number of mobile subscribers was close to 1.2 billion at the end of 2002, while the number of fixed lines was slightly below 1.1 billion. The year 2002 therefore established at worldwide level what had already been observed for an increasing number of countries during the previous few years: mobile telecommunications is the most widespread access tool for telecommunications services. The mobile telecommunications industry has acquired as many users in some twenty years worldwide which took the fixed line telecommunications industry more than 120 years to achieve.

The timely and efficient supply of mobile telecommunication services has had a substantial impact on the economy, which also explains the extensive public interest in this industry. The actions of the industry regulator are of crucial importance for this. For instance, a study on the US market shows that the regulatory delay in licensing mobile telecommunications gave the US consumers welfare losses in the range of $24–50 billion a year.[1]

As will be shown in this book, two factors have determined the extraordinary rapid development of this industry: *technological progress* and

[1] This figure is quoted from Hausman (1997). However also other studies such as Rohlfs, Jackson and Kelley (1991) find such orders of magnitudes.

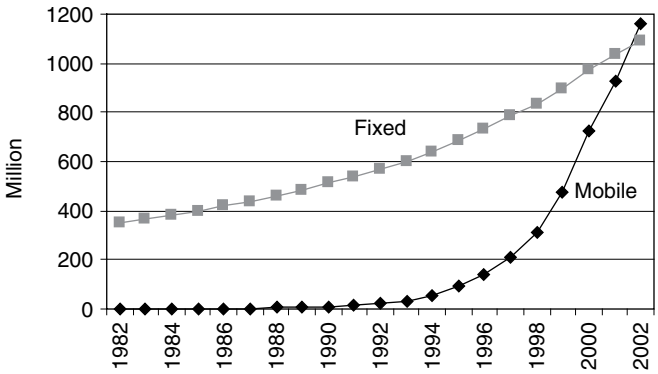

Figure 1.1 *The evolution of the worldwide number of mobile and fixed telecommunications lines, 1982–2002*
Source: ITU data.

regulation. The mobile telecommunications industry as it is known today – i.e. using radio waves instead of wires to connect users – is a relatively young industry. However, its basic technological concepts actually date back to the second half of the nineteenth century, when the German scientist Heinrich Rudolf Hertz demonstrated (in 1888) that an electric spark of sufficient intensity at the emitting end could be captured by an appropriately designed receiver and induce action at a distance. The first mobile telecommunications systems were based on the same principles as radio or television broadcasting, by which all conversations could be heard by everybody. These systems had very limited capacity and used the electromagnetic radio spectrum, whose usable portion is only very limited, in a very inefficient way. Significant progress in using the spectrum more efficiently and ensuring privacy in conversations were made with the development of the 'cellular' concept after the Second World War. However it took until the 1970s for the progress in semiconductor technology to allow the construction of cellular mobile networks for commercial use. Analogue technology cellular systems were introduced first at the beginning of the 1980s. The breakthrough for a mass market for mobile telephony occurred only in the 1990s with the advent of digital technology. The scarcity of radio frequencies, necessary for transmission between the user's handset and base stations, has since then constituted the bottleneck for the development of the industry. As we have seen, the early analogue technology used the allocated radio frequency spectrum in a relatively inefficient manner so only a relatively small number of subscribers could be connected, who used the system mainly for business purposes. The

introduction of digital technology led to a breakthrough in performance, capacity and quality of mobile telecommunications. Digital technology, such as the European standard, GSM, made better use of the radio spectrum than analogue technology did and could therefore accommodate more subscribers. Lower unit costs could be achieved by spreading fixed costs over more subscribers.

Regulatory reform is the other driving force behind the spreading of mobile telecommunications. Because of the radio spectrum constraint, the industry is structurally considered as an oligopoly and the development of the industry crucially depends on pre-entry regulation. In emerging industries, characterised by significant technological progress, there is usually little consensus on the optimum policies concerning the development of the sector. Among other issues, the debate focuses on how and when entry should be promoted and whether technology standards should be imposed centrally or selected by the market forces in a decentralised way. Because of the lack of consensus, governments have taken different policy options, and often change directions as experience accumulates.

The effects of entry in the cellular mobile industry are particularly interesting to analyse. Radio spectrum is the scarce resource to be assigned and constitutes the entry barrier for the firms. However, technological progress permits greater efficiency in spectrum usage and thus potential for accommodating more firms. Governments throughout the world have also taken quite different options regarding the timing and the number of entry licences. This provides interesting data for assessing the effects of licensing on the evolution of the industry.

Such pre-entry regulation in mobile telecommunications has various dimensions. First, the policy maker needs to decide whether to set a single national (or international) standard, or whether to allow multiple technological systems to compete. Second, the policy maker has to decide how many firms a licence will be granted. This also involves an important decision with respect to the timing of first and additional licences. Third, the government needs to decide how to grant licences. In the early days of mobile telecommunications, licences were often granted on a first-come-first-served basis. With the introduction of the cellular technology, the first licences were frequently granted by default to the incumbent fixed operators. Additional licences were initially granted through an administrative tender procedure (lotteries, or 'beauty contests') and then more and more through auctions. This evolution has greatly changed the nature of the firms in the market and their competitive behaviour.

Economic theory can give guidance on these issues, but the propositions of traditional textbook economics are complicated by the fact that mobile telecommunications is a network industry. For instance, in markets

without network effects, it seems to be unambiguously desirable to allow multiple competing technological systems. In contrast, in markets with network externalities there are both advantages and disadvantages to having multiple systems rather than a single standard. The presence of (strong) network externalities typically leads to 'tipping' markets, where the winning technology takes the whole market. Should the government intervene in this race by imposing a single standard? Or should the markets decide themselves on which standard will eventually 'win'? The theoretical literature does not provide an unambiguous answer to these questions.[2]

There is also the question to which extent network externalities are in fact present in cellular telecommunications markets. The main sources of network externalities arise from the fact that mobile users can use their handset only within the areas that support their technological system. Thus, depending on the mobility of consumers, network externalities are local, national, or even international in scope. In addition to reducing consumer switching costs and creating 'roaming' possibilities, the presence of a single technological system also has the traditional advantage of exploiting economies of scale in the manufacture of equipment. Various incompatible technological systems have been developed in the cellular mobile telecommunications industry (most of them with the support of leading countries). Each system is subject to network externalities in that consumers value a system more the more users adopt it. The relevant policy question is whether governments should impose a single standard, or whether the markets should select a winning standard in a decentralised way. Advantages of mandatory standards are that potential network externalities can be realised faster, and that users' technological uncertainty is reduced. Advantages from a decentralised approach are that there may be less a risk of being 'locked in' with inferior technologies and that incentives for innovation to better systems are preserved. Yet a counter-argument is that also the decentralised, market-based, approach may lead to lock-in with inefficient technologies. Despite the extensive theoretical literature, there exists little empirical work that compares the effect of imposing standards on the diffusion of a new technology with the effect of allowing multiple systems to compete. Again, the cellular mobile telecommunications industry offers an interesting opportunity to make such a comparison, since countries have followed quite different and changing policies regarding standards. While chapter 2 gives a general overview of the main issues affecting the mobile telecommunications service industry, chapter 3 is an extensive description of the evolution of the mobile telecommunications industry looking at representative countries. The aim is

[2] See, for instance, Katz and Shapiro (1994) and Shapiro and Varian (1999).

to highlight the importance of country-specific effects, especially at the beginning of the industry. These country-specific effects tend to peter out as the industry progresses. Chapter 4 provides answers to questions of the role of different regulatory policies on the diffusion of cellular mobile telecommunications, relying on quantitative methods and using a world-wide data set.

1.2 Business strategies for firms

One of the main features of a mobile telecommunications network is to provide *coverage*. The fact that a user can utilise a mobile phone over a very large portion of the territory distinguishes it from the fixed network. This coverage can be provided by only a limited number of firms. The radio spectrum bottleneck acts as barrier to entry and makes the industry intrinsically oligopolistic. The question arises which type of strategies firms are able to pursue in such an environment concerning pricing and product positioning. For instance, there may be scope for vertical product differentiation by providing different levels of coverage. However, differentiation in coverage seems to be possible to only a limited extent, mostly during the early years of the life cycle of the industry, when firms have to spread network build-out over time for cost reasons, but in the longer term firms typically have regulatory obligations to provide full coverage. This means that there is little scope for relaxing price competition through product differentiation in terms of coverage. But when differentiation is possible, studies shown that price competition is relaxed. Empirical studies also show that price competition is of the Cournot type, i.e. with price above marginal cost and decreasing with the number of firms in the market.

Pricing of mobile telecommunications services is multidimensional and hence complex, both at the wholesale and the retail level. Retail pricing decisions concern mainly services such as subscription, on-net and off-net calls. Wholesale pricing also include interconnection pricing among networks. Theory provides limited guidance, as the economic literature still has to explore many aspects of pricing in network industries. The market power of individual firms may be exerted to a different degree at each level. It may thus be important from a social welfare point of view to check abuse of market power through '*ex ante*' regulation – i.e. through measures that limit damaging behaviour before it occurs. There is a consensus among the policy makers that such '*ex ante*' regulation, if necessary at all, should be as light as possible. This implies that such regulation should be much lighter in mobile telecommunications than in fixed telecommunications, where 'natural monopoly' positions seem to be much more entrenched.

Regulators took some time to appreciate that cost allocation mechanisms could be profoundly different between fixed and mobile networks. While fixed network infrastructure used to be based on plant and equipment that from an accounting point of view had been depreciated, mobile network infrastructure was typically new and thus carried high depreciation charges in cost accounting. This, for instance, led to regulated interconnection prices that were favourable to mobile telecommunications firms. Cost allocation mechanisms are important when it comes to establishing other aspects of interfirm compensations and how these are transferred to the users. There are two principles: calling party pays (CPP) and receiving party pays (RPP). Although from a theoretical point of view RPP seems to have better characteristics for ensuring allocative efficiency, CPP has been the overwhelming success in terms of worldwide diffusion. Only a few countries, in particular the USA, actually have RPP in place, and for legacy reasons rather than for choice. CPP allows firms to exercise market power in call termination. The favourable interconnection arrangements with CPP provided the mobile telecommunications industry with the financial resources for subsidising the acquisition of customers, and this may account for a substantial part of the rapid growth in the mobile telecommunications subscriber base. Regulatory attempts are underway to fence in the market power mobile telecommunications firms have on traffic termination. Similar considerations apply for international 'roaming', where there are actually elements of RPP but where firms are nevertheless able to exploit the lack of information on the customer side. In any case, the evolution of overall mobile telecommunications service pricing shows a general trend towards more competitive pricing, but there are still some large areas where this does not apply. These issues are addressed in detail in chapter 5, which sets a framework for the business strategies concerning product positioning and pricing. Particular attention is devoted to market segments where market power can be exercised more easily.

1.3 Radio spectrum availability as a key determinant for market structure

Radio spectrum, the key input for the supply of mobile telecommunications services, is a public good, but its use is exclusive when employed for mobile telecommunications services. Its allocation thus needs to be regulated. Other services such as broadcasting compete for the allocation of spectrum and hence only a limited portion of the spectrum is available for mobile telecommunications services.[3] This combined with the high sunk

[3] The technical properties of the radio spectrum and the technical description of mobile telecommunications are discussed in more detail in the appendix.

costs for the set up of mobile telecommunications networks leads to the consequence that the market can support only few firms. The frequency assignment mechanism to firms is important for two reasons: first, radio frequencies are a scarce resource for exclusive use; second, radio frequencies provide a potential for oligopoly rents. For these reasons, the spectrum assignment method is very important and can be divided into two major categories: administrative methods, such as 'beauty contests', and market-based methods, such as auctions. There are refinements for each category, but the main difference boils down to the role of information retained by the government. With administrative methods the government plays an active part in the assignment and advocates a central role in the development of the industry. With a market-based method the government prefers to compare itself with a referee, setting the framework and letting firms decide on the implementation of measures for the development of the market. For instance, the auction mechanism is based on the belief that the market has sufficient capability for self-selection to award spectrum to the firms that make the most efficient use of it, and that this is in the public interest. Chapter 6 surveys the different assignment mechanisms, presenting their advantages and disadvantages. The experiences of selected countries are documented in some detail to illustrate these points. The most important episode in this respect is the assignment of so-called 'third generation' (3G) licences in Europe. This has shown that auctions in general deliver much higher receipts to the governments than do administrative methods. Moreover, the design of auctions, in particular with the aim of avoiding collusion, is of utmost importance in generating large receipts. However, serious doubts have arisen on whether bidding agents are really better able than governments in assessing market prospects.

Although entry into the mobile telecommunications market is regulated, there is the question whether the industry is a 'natural oligopoly'. If spectrum were not a scarce resource, other factors, such as sunk costs or scope for vertical product differentiation, could set in as determinants of market structure. The historically observed evolution of market structure in the mobile telecommunications industry is from higher to lower levels of concentration. In most countries, the industry has evolved from a monopoly to an oligopoly with three or more firms. Waves of generations of technology have typically been a trigger for additional entry, as newer generations of technology with more efficient use of radio spectrum permitted the entry of more firms. This entry has been sequential, and the profitability of the industry has declined, with new entrants being less profitable than long-established firms. The question now arises of whether entry has led the industry to the zero profit level. This could be indicated by the observed exit or attempts to merge of late entrants, as being noted in

some countries. This is particularly relevant in the forthcoming market for 3G mobile services in Europe, where a new design of market structure has taken place. Governments decided for simultaneous entry of a larger number of firms than for 2G (second-generation) mobile services, but with apparently little assessment of whether the new market would support such a large number of firms. Moreover, there has been a tendency to privilege auctions as the assignment method. It has turned out that with auctions there is a tendency to increase the number of firms in the industry, and with individual firms paying more than with other assignment methods. Chapter 7 develops a benchmark model that illustrates the interplay between sunk costs, such as licence fees, and market structure. It suggests that 'overbidding' of licence fees may occur, at the expense of forsaking the market structure envisaged by the policy maker or of collusion at the post-entry stage in the market. The model's predictions are compared with the outcomes from 3G licensing in Europe and subsequent events. Evidence of exit of firms and calls for relaxation of licence conditions suggest that overbidding had taken place: even in cases with zero licence fees the government has apparently allowed for too much entry. This may suggest that the industry has arrived at a point where spectrum is no longer a constraint. Entry may not even need regulation any more. If this were to be the case, it would mark the emancipation of the industry from the spectrum bottleneck.

2 Stylised features of the mobile telecommunications industry

2.1 Introduction

Mobile telecommunications use radio waves,[1] instead of wires, to connect users. Though the origins of wireless communications may be traced back to the second half of the nineteenth century, the earliest applications for mobile communications date back to the 1920s. After the Second World War, when the civilian use of wireless telecommunications resumed, several industrialised countries independently developed mobile telecommunications systems. These, however, suffered of a series of technical limitations that hampered their widespread use. Only during the 1980s did these problems begin to be surmounted, with the diffusion of cellular mobile telecommunications technology as it is known today. To fully appreciate the technological challenges mobile telecommunications had to surmount to become a widely spread technology, it is useful to briefly sketch the history of the technology in the context of the working principle of wireless communications. This chapter outlines the main driving forces of the mobile telecommunications industry and how they shape the evolution of the sector and gives some hints on the prospects for the future of the sector. The key issues will be dealt with in more detail in subsequent chapters. This chapter is organised as follows. Section 2.2 presents a brief history of the technological developments in the mobile telecommunications industry. Section 2.3 provides some notions of the different technologies available. Section 2.4 illustrates some of the main user trends in this fast-growing industry, while section 2.5 looks at the revenue side. Section 2.6 takes a closer look at the cost side, which proves to be very important in driving penetration of mobile telecommunications: even

[1] Radio waves are a natural resource and only a small part of the total electromagnetic spectrum is suitable for radio transmission. The measurement unit is Hertz (Hz) which indicates the cycle per second. For more technical details, the reader is referred to the appendix.

though most of the cost elements are declining in this industry, some are increasing and could be of crucial importance. Section 2.7 discusses the main issues concerning regulation, both the pre- and post-entry stage. Section 2.8 draws some brief condusions.

2.2 Some technology history

2.2.1 Mobile telecommunications before the cellular era

The first attempts at wireless communications

The origins[2] of wireless communications may be traced back to the German scientist Heinrich Rudolf Hertz, who demonstrated in 1888 that an electric spark of sufficient intensity at the emitting end could be captured by an appropriately designed receiver and induce 'action at a distance'. This transmission via the 'ether' challenged the classical notions of physics. Whereas Hertz's experiments spanned just a few metres, it was the Italian scientist Guglielmo Marconi who constructed a 'radio' that transmitted waves over increasing distance: In 1895, he transmitted signals over a distance of 2.5 km, in 1899 over the English Channel and in 1900 over more than 300 km.[3] Marconi's greatest challenge was to confute the conventional belief that radio waves propagated only linearly and therefore would be unable to follow the curved surface of the earth. In 1901, Marconi established the first wireless transmission over the Atlantic, spanning over 3500 km from Cornwall to Newfoundland. Maritime applications become the dominant market for wireless, even though only large and expensive ships could carry the wireless equipment and justify the cost.

At the beginning, only gross pulses of energy could be transmitted, and communications was limited to Morse code. Technological improvements, in particular the refinements in radio communication technology such as amplitude modulation (AM)[4] and the invention of the thermo-ionic valve, led to the possibility of transmission of speech and music. However, wireless equipment was a low-volume and high-cost market. Before the start of the First World War there were some 2000–3000 wireless in use in the entire world, most of them in Britain. At the outbreak of the war in Europe the development of wireless was intensified, again mostly for

[2] Historical accounts of the industry can be found in Calhoun (1988), Mehrotra (1994) and Garrard (1998), who also refer to primary sources.
[3] In 1896, Marconi offered his wireless system to the Italian government, but he never received a reply and eventually he decided to emigrate to England. There, he met Sir William Preece, the chief engineer of the telegraph office, who provided Marconi with the funds for an experimental site at Lavernock in Wales UK.
[4] Undertaken the first time by Reginald Fessenden in 1905, with AM information transmitted by varying the amplitude of radio waves.