Chapter 1

Introduction

1.1 Two-Way Interconnection

This work is concerned with the economics of networks characterized by two-way interconnection. The term interconnection is widely used in the literature on network industries to describe the linking of different networks. Specifically, for telecommunications networks, the European Community's interconnection directive 97/33/EC defines interconnection in the following way:

'Interconnection' means the physical and logical linking of telecommunications networks used by the same or a different organization in order to allow the users of one organization to communicate with users of the same or another organization, or to access services provided by another organization.

The problem with this formulation is that it subsumes two quite different meanings in one word. The usual understanding of interconnection is the one given in the first part of the definition, i.e. the physical and logical linking of telecommunications networks used by the same or a different organization in order to allow the users of one organization to communicate with users of the same or another organization. The second part of the definition also allows the term interconnection to refer to the linking of telecommunications networks in order to allow the users of one organization to access services provided by another organization. This, however, is something rather different. In order to avoid confusion, the type of interconnection addressed in the second part of the definition is usually called access to a network,
or, alternatively, one-way interconnection. The first part of the definition is then what is often referred to as two-way interconnection, and this is also the interpretation of the term interconnection we use in this work.

We do not intend to study the technical problems of interconnection, nor do we address the legal obligations which are partly responsible for the existence of interconnection of networks. We start with the observation that networks, especially telecommunications networks are interconnected, and then we ask what are the economic implications of this for the network industry, the single networks, and the end users of the networks. This is the meaning of the title of this book, ‘The economics of two-way interconnection’.

1.2 Telecommunications

Why is the market for telecommunications an interesting object of study? What makes it so different from the market for, say, apples? A short answer would be that telecommunications involves all the characteristics which are typical for so-called network products. This is not the whole truth, however, and the question deserves a more detailed answer. In the following paragraphs, we provide some keywords which serve to make clear some of the important special features of the market for telecommunications. We start with a simple situation: Imagine Anna wants to talk to Bob, who lives in another town. The simplest way to do this is to make a phone call to Bob. What are the characteristics of the telecommunications market Anna encounters by doing so?

- **Complementarities:** Anna could go and search for a public phone booth, but if she wants to call Bob regularly, she might find it useful to buy a telephone of her own. However, owning a telephone is not enough. Additionally, Anna’s telephone must be connected to an active telephone line which makes access to the public telephone network possible. Anna’s telephone is only useful in connection with the telecommunications services provided by the operator of the public network. Without a telephone, on the other hand, these services are themselves useless. The telephone and the services provided are *complements*.

- **Network externalities:** Even if Anna owns a telephone and is connected to a network, this will only enable her to talk to Bob, if Bob is also
connected and has a telephone. This is true for any person Anna might want to call. Hence the value of the phone and the service to Anna depends strongly on the number of other users which are connected. If Bob buys a phone and subscribes to the network, this increases Anna’s willingness to pay for her own services, a positive externality which is called network externality.

- **Call externalities**: Anna calls Bob because she likes to talk to him. When Bob’s telephone rings, he will most probably answer it, even if he does not know who the caller is. The reason he answers the phone is that his expected utility of doing so is positive. Bob might receive nuisance calls now and then, but usually, and on average, he benefits from being called. The utility he receives from Anna’s call is a positive externality known as the *call externality*.

- **Economies of scale**: If Bob lives in the same country as Anna, and if both use a fixed-line telephone for their call, it is most likely that they are connected to the same network. The reason for this is that fixed-line network operators within a country are usually monopolists. Establishing a fixed-line network involves huge sunk costs, basically because it requires wiring the whole country with copper lines and physically connecting each household to the nearest switch of the network via the ‘last mile’. Once the network is set up, however, delivering calls through it generates only marginal costs which are quite small. These significant *economies of scale* have led to the view that the telecommunications industry is a ‘natural monopoly’, and governments have usually licensed only a single company to provide telecommunications services.

- **Interconnection**: If Anna and Bob live in different countries, or if they use a mobile phone for their call, it is quite possible that they will be connected to different networks. The natural monopoly argument does not apply for mobile telecommunications, since it is much easier and less costly to establish a mobile network, where the ‘last mile’ is bridged via electromagnetic waves rather than via copper cables. However, if Anna and Bob are subscribed to different mobile telecommunications networks, they can only conduct a phone call if these networks are interconnected (in the sense of two-way interconnection as explained above). Since different networks, even if they are interconnected, usually com-
pete with each other for customers, this raises additional problems on the supply side.

- **CPP vs. RPP:** A part of the costs of Anna’s call to Bob is borne by the network Bob is connected to. These are particularly the costs of transmitting the call from the point of interconnection to the base station next to Bob’s position, and the costs of terminating the call at Bob’s mobile phone. There are basically two possibilities for Bob’s network to be compensated for these costs. It could either charge Bob for receiving a call on his mobile phone, or it could charge Anna’s network, which will then collect (part of) this charge from Anna via increased retail prices.\(^1\) The first payment system is called the ‘Receiving Party Pays’ (RPP) system, and the second one the ‘Calling Party Pays’ (CPP) system. CPP is the de facto standard in Europe, while RPP is used in the USA, Canada, and Hong Kong. In this work we concentrate exclusively on the CPP system.

- **Termination-based price discrimination:** If Anna calls Bob on her mobile phone, the price she pays for a call-minute will most likely depend on whether Bob is connected to the same or a different network. In the former case, the call does not leave their network, it is called an on-net call. In the latter case, the call is transferred from Anna’s to Bob’s network at the point of interconnection, a switch where calls are routed into the network where they terminate. In this case we speak of an off-net call. If a network’s price for on-net calls differs from the price for off-net calls, the network engages in termination-based price discrimination. Interestingly, this type of price discrimination introduces another kind of externalities:

- **Tariff-mediated network externalities:** Assume for the moment, that the price of on-net calls is lower than the price of off-net calls. Assume also, that the market is covered, i.e. each consumer is subscribed to some network. Under these circumstances the standard network externalities discussed above are exhausted, the sum of the networks’ customers cannot grow any more. Nevertheless, Anna benefits if the market share of her network increases. The reason is that she can reach more of her

\(^1\)In principle Bob’s network could also charge Anna directly, but this possibility is usually ruled out for technical and/or legal reasons, because it would require Bob’s network to be able to bill virtually any person who might want to call one of its customers.
calling partners on-net, where it is cheaper to call. Hence Anna benefits (in expectation) whenever some customer of another network switches to her network. This constitutes a positive externality. However, since it does not work directly, but is mediated through the discriminating tariffs, it is called a *tariff-mediated network externality*. It should be mentioned that some economists refuse to call this effect an externality at all, since it relies on the price system to work. However, this is mainly a matter of how strict one defines the term 'externality'.

- **Balanced calling patterns**: Note that the presence of tariff-mediated network externalities as explained above relies on the implicit assumption that there is a positive probability for Anna to call any other customer. To see this, imagine that Anna exclusively calls Bob, and nobody else. Then, as long as Bob is subscribed to the same network as Anna, she does not care at all about her network’s market share. The tariff-mediated network externalities exist only between Anna and her calling partners. If one assumes, as is regularly done in the literature, that Anna is equally likely to call any other customer, then we speak of *balanced calling patterns*. If calling patterns are balanced, the percentage of all calls terminating on some network is equal to the percentage of all calls originating on this network, and both are equal to this network’s market share. This makes calculations much easier in our models, but the question remains if the assumption of balanced calling patterns is really well-founded.

- **Nonlinear pricing**: Anna and Bob do not only pay for their phone calls, they are also likely to pay a fixed monthly fee for being connected. This means that the pricing scheme is **nonlinear**.¹ If a fixed fee is combined with a per-minute price for calls, we speak of a *two-part tariff*. A network using a two-part tariff has two distinct instruments to compete for market share and generate profit. Note that the fixed fee and the price for a call affect demand rather differently. While demand for subscription in general depends on both the fixed fee and the per-minute charge, a subscribed customer’s call volume is independent of the fixed fee, at least in the absence of wealth effects.

¹Total payment as a function of minutes called is still an *affine* linear function, but the term *linear pricing* solely denotes the case where this function is strictly linear, i.e. where there is no fixed fee.
• **Access charges:** If Anna and Bob are subscribed to different networks, then, as mentioned above, the call originating from Anna’s mobile phone is switched to Bob’s network at some point of interconnection. From there it is transmitted to the base station next to Bob, and the ‘last mile’ to Bob’s mobile phone, where the call terminates, is bridged via electromagnetic waves. This means that parts of the total marginal costs of the phone call are borne by Bob’s network. Under CPP, Bob’s network will charge Anna’s network a per-minute price for terminating Anna’s call. This price is sometimes called *termination charge* or *interconnection fee*. Unfortunately, in the academic literature the somewhat misleading term *access charge* is the most commonly used one. While access is something different from interconnection, as discussed in Section 1.1, we will nevertheless, for the sake of continuity, also use the term ‘access charge’ for what should really be called ‘termination charge’.

### 1.3 The Role of the Access Charge

Even if there is a large number of network operators in a telecommunications market, a single network is not subjected to any competition in one particular part of its services, namely in terminating calls to its customers. It is obvious, and seems technically unavoidable, that networks retain a monopoly position with respect to termination of such calls, and since networks can generate profits on incoming calls by raising their access charge above marginal cost, all the regulatory concerns usually associated with monopoly power also arise in these markets.

Should access charges be regulated, and if so, how? Before we address this question within a formal model, we will try to provide some intuition about the role these access charges play for prices, profits, and welfare.

The first thing to notice is that access charges are part of the *perceived marginal costs* of an off-net call. As mentioned earlier, an off-net call generates ‘true’, technical marginal costs for the originating network up to the point where the call is handled over to the rival network. The remaining part of the marginal costs as perceived by the network originating the call are the access charges paid to the network terminating the call. Since these access charges need not be equal to true marginal costs of termination, the
perceived total marginal costs of an off-net call will in general be different from the true total marginal costs.

When setting its prices, a profit maximizing network will equalize marginal revenue and marginal costs. However, it is the *perceived*, not the true marginal costs, it will take into account in this optimization procedure. Hence end-user prices of a network are (partly) determined by perceived marginal costs, and hence by its rivals’ access charges. Particularly, a network will raise its price (for off-net calls, in the case of termination-based price discrimination), if one of the rivals increases its access charge.

Now let us consider what happens if two competing networks are allowed to set their access charges independently, i.e. noncooperatively.

Consider two symmetric networks competing in a covered market. By ‘symmetric’ we mean that the networks are basically identical, they have the same cost structure, pricing method, quality delivered, and so on. Under some additional assumptions, outlined in the following chapters, there will be a symmetric equilibrium, where these networks set the same access charges, offer the same prices, and share the market equally. With balanced calling patterns, the total number of calls originating from a network and terminating on the other network, i.e. the number of off-net calls, will be the same for both networks. Hence the total access charge payments a network makes to its rival is exactly equal to the total access charge payments it receives. These payments cancelling out, it seems as if profits are completely unaffected by the access charge.

This is not the case, however. Indeed, the belief that access charges play no role is widespread,\(^3\) and has been termed the *bill-and-keep fallacy* by Laffont and Tirole (2000). As we have explained above, access charges have no direct, but an indirect effect on profits through their influence on end-user prices. If a network, let us call it network A, unilaterally increases its access charge, its rivals’ average end-user price will go up, and as a consequence, indirect utility of the customers of A’s rivals goes down. Since customers compare the net utilities they receive from the available networks in their subscription decision, they will tend to switch to network A, increasing A’s market share and profit.

\(^3\)Armstrong (1998) mentions a paper by the New Zealand Ministry of Commerce supporting this wrong intuition.
The same is true, however, for the rival network. In the end, both networks will increase their access charges, and this may well raise call prices above the monopoly level. This is detrimental to both consumer and producer surplus, and hence calls for regulatory intervention. Note that this effect, which has been called the *raise-each-other's-cost effect*, is similar to the *double marginalization problem* arising in vertically structured industries and well known in the theory of industrial organization.

For these reasons, it is commonly agreed that networks should not be allowed to set their access charges noncooperatively. One way of alleviating the double marginalization problem is to impose *reciprocity* of access charges, i.e. to demand that both networks charge the same unit access fee. This can be achieved by a regulator setting an appropriate reciprocal access charge, or by letting the networks freely negotiate over the access charge, subject only to reciprocity. In many OECD countries, interconnection arrangements are indeed handled in the latter way, with regulatory intervention only if negotiations fail. Now, while collusion over retail prices is illegal in general, cooperative agreement on a reciprocal access charge is not only allowed, but often encouraged. This makes sense only if firms are not able to indirectly collude over retail prices by colluding over the access charge. Unfortunately this is by no means obvious.

### 1.4 Literature Overview

#### 1.4.1 Linear Pricing Models

In the second half of the 1990s, serious concerns have been raised in the literature about firms' ability to use a cooperatively determined access charge as a collusion device (see e.g. Brennan (1997)). As noted already by Katz et al. (1995), networks have an incentive to agree on a high (above marginal cost) reciprocal access charge in order to achieve high end user prices. Together with the confirming results from the first explicit models (see below for details on this literature), this has led many researchers to adopt the view that collusion in the retail market is associated with *high* access charges. This view was only slightly clouded by subsequent opposite results arising from refinements of the basic models, which tried to eliminate some of the less realistic assumptions of these models.
The first to show the negative welfare effects of cooperatively determined access charges within an explicit model were Armstrong (1998), Laffont et al. (1998a) — henceforth LRTa — and Carter and Wright (1999). They employ models where two networks are horizontally differentiated in the Hotelling style and compete for customers in linear, nondiscriminating prices. The model of LRTa is by now widely accepted as the “standard model” of two-way interconnection, and most of the subsequent literature uses this model as a starting point. Basic assumptions of LRTa’s model include that consumers do not benefit from receiving calls and that calling patterns are balanced. All these authors conclude that the negotiated access charge may be used as a collusive device and will definitely exceed the marginal cost of access.

1.4.2 Nonlinear Pricing Models

If networks may compete in nonlinear prices, e.g. two-part tariffs, this result does no longer hold. As LRTa show, equilibrium profits are independent of the access charge, leaving networks indifferent about the price of interconnection. The intuition is that although usage fees still increase with the access charge, networks can counterbalance the negative impact on market share by lowering the fixed fee. Thus competition remains strong, and the access charge looses its collusive function.

Dessein (2003) studies a model where consumers differ in volume demand or subscription demand. He shows that introducing heterogeneity in volume demand leaves the neutrality of the access charge unaffected. This result is also supported by Hahn (2004). If demand for subscription is elastic, however, some consumers may choose not to subscribe in equilibrium. As Dessein (2003) and Schiff (2002) show, this leads networks to prefer an access charge below marginal cost. The reason for this is the emergence of positive network externalities in the absence of full participation.

1.4.3 Termination-Based Price Discrimination

The mentioned models do best describe local fixed-line telecommunications networks. With the rise of mobile telecommunications, however, the practice of termination-based price discrimination became apparent. In mobile networks it is commonly observed that different prices are charged for calls
terminating in different networks. Termination-based price discrimination was already studied by Economides et al. (1996). However, their results differ substantially from the results discussed below, since they assume that subscription decisions are made before prices are set, which renders market shares effectively exogenous.

A seminal paper introducing price discrimination into the models mentioned above is Laffont et al. (1998b), henceforth referred to as LRTb\(^4\). Among other results they show that with linear pricing, the collusive role of the access charge is reduced by the possibility of price discrimination. The reason is that similar to the case of two-part tariffs above, a higher access charge is reflected in a higher off-net price, but the building of market share is not necessarily linked to an increase in the access deficit, since customers can be attracted by lowering the on-net price. However, as opposed to the nondiscriminatory, nonlinear pricing case, the collusive role of a high access charge is not completely removed. Proposition 2 of LRTb states that the access charge still \textit{locally} acts as a collusion device, which means that profits increase locally, if the access charge is increased above marginal cost.

As in the nondiscriminatory case, the corresponding result for nonlinear prices is quite different. Gans and King (2001) demonstrate that networks competing in two-part tariffs with discriminating call prices will negotiate a very \textit{low} (below marginal cost) access charge in order to soften competition. They also conclude that the widespread bill-and-keep arrangements, corresponding to a zero access charge, may be undesirable from the consumers’ perspective. As Cherdron (2001) notes, however, their result, predicting off-net prices below on-net prices, is somewhat at odds with what can be observed in existing mobile networks.

Also other authors have asked if bill-and-keep arrangements, which are usually argued to save transaction costs, are actually anticompetitive. A natural benchmark against which the welfare effects of such an agreement can be evaluated is cost-based access pricing, which sets access charges equal to marginal cost, corresponding to conventional regulatory wisdom. As mentioned above, Gans and King (2001) favor cost-based access charges, arguing that bill-and-keep arrangements may be used to soften competition. An opposing position is taken by Cambini and Valletti (2003), who demonstrate

\footnote{A summary of the results of LRTa and LRTb is given in Laffont et al. (1997).}
that bill-and-keep arrangements may be beneficial due to a positive impact on investments in quality prior to the competition stage.

Given the important characteristics of mobile telecommunications markets outlined above, it is surprising that the literature completely lacks a model of a caller-pays system incorporating nonlinear pricing and termination-based price discrimination as well as call externalities. Laffont et al. (1998b), Gans and King (2001), and Cambini and Valletti (2003) study the case of nonlinear discriminatory pricing, but without call externalities. Kim and Lim (2001), DeGraba (2003), and Jeon et al. (2004) take into account the call externality, but they concentrate on receiver-pays systems (where the importance of call externalities is more obvious). Hahn’s (2003) model has nonlinear pricing and call externalities but studies a monopolistic network. Finally, Armstrong’s (2002) extensive survey includes a small study of nonlinear pricing in the presence of call externalities, but without price discrimination.

### 1.5 Introducing Call Externalities

Summarizing the above, while under nonlinear pricing networks are either indifferent about the access charge or prefer an access charge below marginal cost, the work concerned with the linear pricing case unanimously suggests that networks will negotiate a high access charge to maximize joint profits. Subsequently, we will show that actually the opposite might be the outcome of network competition in linear prices, and networks might well make use of a reciprocal access charge below marginal cost. This result may look similar to the one of Gans and King (2001), but there is an important difference. While their result has been criticized for being out of line with observed price structures, this does not apply to our findings, at least in the linear pricing case. Access might be sold at a discount, but off-net prices still exceed on-net prices in equilibrium. Moreover, there turns out to be little scope for regulatory intervention against bill-and-keep arrangements. These arrangements might result from collusion, but then they are also welfare improving compared with cost-based access pricing.

However, for competition in two-part tariffs, the Gans and King (2001) result is confirmed if receivers’ utility is taken into account. The negotiated access
charge is always below marginal cost, and off-net calls are cheaper than on-net calls.

All of the papers discussed in the introduction share the basic assumption that a call generates utility only for the caller and not for the receiver. In this work we divert from this assumption by introducing call externalities. The obvious point that a call generates utility also for the receiver has been recognized\(^5\), but nonetheless widely neglected in the literature. Only recently, Kim and Lim (2001), and Jeon et al. (2004) have come up with similar models incorporating a call externality. However, they study a RPP system, where both the caller and the receiver of a call are charged. Note that the receiver of a phone call incurs the opportunity costs of the time the call takes. Hence he must get some strictly positive utility from a call, otherwise he would not answer the call. On the other hand it might be argued that at least on average the utility of the receiver will be smaller than the utility of the caller. Whatever the “real” average magnitude of receivers’ utility, neglecting it is likely to introduce a relevant distortion in the analysis of network competition.

First, however, it can be seen that under nondiscriminatory pricing the analysis of competition remains unchanged\(^6\). It is clear that volume demand is independent of any call externality. Obviously, nondiscriminating prices also make the subscription decision independent of receivers’ utility. Hence neither subscription nor volume demand or profits are influenced by the level of passive utility. This means that the results derived from the standard model of nondiscriminatory pricing discussed above carry over to the extension we study here. The only deviation from LRTa’s model arises in the judgement of welfare implications. Indeed, neglecting the call externality underestimates social welfare. To implement the social optimum, the price of a call would have to be below marginal cost.

Volume demand stays of course independent of the call externality also with termination-based price discrimination, but the subscription decision is influenced if on-net prices differ from off-net prices. This is because the utility from receiving calls contributes to the positive network externality under on-net prices (say) below off-net prices. An increase in a network’s market share

\(^{5}\)DeGraba (2003) suggests that the total utility generated by a call is shared equally between the calling parties. See also the discussion in Hahn (2003).

\(^{6}\)See also the discussion in Schiff (2001).
raises the number of calls received by (and hence benefits the) subscribers of this network. In their subscription decision, consumers compare the net utilities they receive from joining either network. If a network raises its off-net price, this has two effects. First, the net utility of this network's customers decreases, and second, since these customers' demand for off-net calls falls, also the rival network's customers suffer, because they less frequently enjoy the benefit of being called. This second effect lowers customers' incentives to switch to the rival network. As the access charge, the call externality is reflected in equilibrium prices, which determine profits. Indeed, if the utility of receiving calls is sufficiently high, the second effect explained above becomes so strong that networks will prefer an access discount in order to keep the resulting off-net prices below the monopoly price.

This analysis rests on the assumption that profits are directly determined only by prices. Note, however, that in the case of two-part tariffs profits also depend on the fixed fee. As mentioned above, this has a deep impact on the nature of competition. The case of termination-based price discrimination with two-part tariffs is analyzed in chapter 5.2 of Jeon et al. (2004). Although their work is devoted to the RPP system, they include a short study of their model in the absence of reception charge, which of course coincides with a caller pays system. Interestingly, they show that if receivers' utility is high enough (equal to callers' utility), then for any given level of the access charge, the price for off-net calls in a symmetric equilibrium becomes infinite, resulting in connectivity breakdown. The intuition for this is the following. Any off-net call made generates utility for the caller and the receiver. However, since only the caller pays for the call, if receivers' utility is high, net surplus is higher for the receiver than for the caller. This means that while raising the off-net price may decrease the direct profit from off-net calls, at the same time it makes the own network more attractive, resulting in an increase in market share. The total effect on profit becomes positive, if receivers' utility is high. Furthermore, if receivers' utility is high enough, the total effect on profit is positive regardless of the level of the off-net price. This, of course, means that the only equilibrium has an infinite off-net price.

We conclude that the introduction of call externalities has a strong impact on the outcome of competition in the case of termination-based price discrimination. This is the case we study in Chapters 3 and 4 of this work.
1.6 Outline

In Chapter 2, as a starting point, we describe the standard model of LRTb used in the majority of the literature on network competition. In Chapter 3 we extend the standard model with linear pricing by call externalities. We show that some of the conventional wisdom on the collusive role of the access charge is overturned under this extension. The impact of call externalities on competition in the standard model with two-part tariffs is studied in Chapter 4. Here we argue that the traditional reasons put forward against the access pricing practice known as bill-and-keep turn out to be ill-founded once call externalities are taken into account. Chapter 5 departs from the duopoly assumption and studies the case of three or more competing networks. We show that the determination of market shares calls for a dynamic setting, and that this raises several serious problems like nonexistence or multiplicity of stable equilibria, or nonconvergence of market shares even for fixed prices. In Chapter 6 we introduce a local interaction structure between agents, thereby giving up the standard but unrealistic assumption of balanced calling patterns. For the usually observed case of on-net prices below off-net prices we show that this typically generates a multitude of consumer equilibria, which creates severe problems for the prediction of market shares.