# Essays on Heterogeneity in Economic Networks

Inauguraldissertation zur Erlangung des akademischen Grades
eines Doktors der Wirtschaftswissenschaften
der Universität Mannheim

Malte Cherdron

Vorgelegt im Oktober 2002

Referent: Prof. Konrad Stahl, Ph.D., Mannheim
Korreferent: Prof. Patrick Rey, Toulouse
Dekan: Prof. Dr. Christoph Buchheim

Tag der mündlichen Prüfung: 18. Februar 2003

# Contents

# Chapter 1

# Introduction

This dissertation, written at the *Graduiertenkolleg "Allokation auf Finanz- und Gütermärkten"* at Mannheim, contains three analyses of different types of economic networks with heterogeneous players. In each model, a specific form of heterogeneity is introduced which interacts with the network effects at the core of the analysis, extending or overturning some key results in the literature.

The first model, on compatibility in a durable-goods monopoly, describes a network of users of a good who benefit from each other using an identical product. This is the kind of non-physical economic network based on positive consumption externalities analysed by Katz and Shapiro (1985), Farrell and Saloner (1985) and many other authors since. The effects in this class of models are particularly relevant for markets for computer software (where the ability to exchange files with other users is an important component of the value of the product), as well as for durable goods consumed with less-durable complements (whose availability increases in the size of the market). The focus in the present model is on the effect of consumer heterogeneity in a durable-goods market where standards are chosen by a monopolist. Choi (1994), Waldman (1993) and others have shown how the firm will be tempted ex post to make a new product incompatible with its predecessor in order to reduce the Coasian (1972) competition it faces from its own output sold in earlier periods. When consumers anticipate this, both profits and social welfare are compromised, suggesting a beneficial role of regulation banning incompatibility.

In the model presented here, heterogeneity leads to an interaction between the standard monopoly effect and the firm's incentive to render the existing goods on the market obsolete. In particular, the coexistence of a mass market and a smaller group of high-valuation consumers creates a trade-off for the firm between increasing the size of the network (to which it implicitly sells access in the future), and maximizing short-term monopoly profits. This trade-off is affected when consumers anticipate the future obsolescence of the goods they are buying, because their willingness to pay (and hence, short-term profits) are reduced while the value of network size is not: The threat of obsolescence increases the relative profitability of increasing output to expand the installed base. In contrast with results in the literature, the inefficiency induced by arbitrary incompatibility does not lead to a loss of welfare, but partially compensates the inefficiency of a monopoly. Thus, at least in some cases, banning incompatibility would reduce, not increase, welfare. Under certain conditions, a case can even be made for forcing (or subsidizing) the monopolist to switch to an incompatible new product when the firm would prefer to continue an inferior but widely adopted old standard.

The second model, on interconnection and collusion, is used to analyze the competition for retail customers between telephone networks, in the tradition of the work by Laffont et al. (1998b) and others. An interesting feature of the telecommunications industry, from an industrial-organization point of view, is the fact that the same firms competing with each other in the retail market also provide each other monopolistically with an intermediate good. This is because each network, in order to give its subscribers access to the entire population of telephone users, must have its competitors terminate calls to those consumers who have chosen to be connected by them. Thus, the termination of calls originated on another network is a mutually-provided intermediate input used to produce telephone services, and the price of this input – the interconnection fee, or access charge – has a significant effect on final prices and on the competition for retail customers.

While this might appear to be a more literal, physical type of networks, the driving force behind the model presented here is the creation of the same kind of non-physical, purely economic “network” externalities as in the first. Consumers are taken to be heterogeneous with respect to their calling patterns: Every telephone user belongs to a group whose members he or she calls more often than the rest of the population. In addition, the firms can offer non-linear subscription tariffs discriminating between calls terminating on their own network and elsewhere. These relatively simple assumptions are descriptively realistic, in particular for many mobile-telephony markets. They are also crucial for the final outcome, because they allow the firms to differentiate their otherwise homogeneous products. A markup on interconnection leads to a higher price of calling users on the other network, resulting in a strict preference for each consumer to join the same network as his or her group. Interestingly, this allows for collusion even under non-linear tariffs, as long as prices can differ according to where a call is terminated. The implication for competition policy is that some form of permanent, sector-specific regulation is required even in a mature market to ensure an efficient overall outcome when firms competing in the retail market need to interconnect. This is in marked contrast with legislation in Germany, for instance, where "workable" competition in the telecommunications industry is deemed a sufficient condition for sector-specific regulation to be replaced by standard anti-trust rules. On a positive note, this model suggests that regulation may achieve its goals simply by banning certain forms of price discrimination rather than through a more involved process of setting prices outright.

In the third model, on bargaing and mergers, an industry is formed by a set of firms establishing a network of bilateral trading links. All trade is based on efficient bargaining in the sense that the optimal allocation in the industry is always achieved. The distribution of the resulting surplus, on the other hand, depends on the bargaining power of each player, which is affected by changes in other parts of the trading network even if the player is only indirectly linked to them. In this cooperative setup, which is based on the analysis of Segal (2000), players can merge or split up, increasing or reducing their joint bargaining share without affecting the overall allocation. While this network of bargaining links is quite different from the virtual ones in the first two models based on consumption externalities, the heterogeneity of the players again plays a crucial role.

In particular, the differences in the players' fundamental characteristics allow for a privately profitable merger to have a larger effect on outsiders than on the merging firms, and to compensate the initial effect of a negative shock to an outsider, without relying on any allocative effects. Moreover, the model can be used to demonstrate how the existence of additional, outside players turns pairwise merger decisions into strategic complements or substitutes, leading to an explanation of merger waves which is based entirely on bargaining considerations. In contrast with the first two models, the absence of allocative effects assures that there are no direct policy implications. However, it is straightforward to base an argument about strategic (and inefficient) technology choice or R&D effort on precisely this kind of bargaining model, immediately reintroducing policy issues.

It goes without saying that these three models are a long way from a general theory of the role of heterogeneity in networks. It is not even clear that such a theory would be very helpful at a general level. Nor are the policy implications presented here necessarily robust enough to interpret them as immediate calls for government action. What these models demonstrate, rather, is that differences between the players in an economic model are of particular importance in networks, where the heterogeneity of agents can interact with the externalities between them. *Vive la différence.*

# Chapter 2

# Compatibility and Obsolescence in a Durable-Goods Monopoly with Heterogeneous Consumers

## 2.1 Introduction

This is an analysis of the incentives for a durable-goods monopolist to introduce arbitrary incompatibility between successive generations of its product. When the number of fellow users with a compatible good matters for each consumer (i. e. when there are positive "network" effects), incompatibility reduces the value of the existing installed base, increasing the price consumers are willing to pay to upgrade to a new product. As shown by Choi (1994), Waldman (1993) and others, the firm will indeed tend to use its discretion ex post and choose incompatibility, achieving a form of "planned" obsolescence which in turn can lead to an oversupply of upgrades.[1] Consumers, on the other hand, anticipate that their purchases will be made obsolete in the future, so ex-ante prices fall, offsetting any positive effect on overall profits. Thus, whenever the decision over intergenerational compatibility is made by a monopolist without ex-ante commitment power, the outcome is likely to be incompatibility, inefficient upgrades and a loss of both social welfare and private profits. This seems to suggest that the outcome could be improved by a ban on incompatible upgrades.

[1] Strictly speaking, obsolescence is not planned ex ante in this setting, as in Bulow (1986), but induced ex post when the firm chooses incompatibility. Hence, "arbitrary obsolescence" would be more correct.

The contribution of the present paper is to explain why this result is not robust against the introduction of consumer heterogeneity. In particular, the presence of a large “mass market” of low-valuation consumers leads to an interaction between network externalities and the standard monopoly effect which can reverse both the qualitative results and the policy implications found in earlier work.

In this model, consumer heterogeneity leads to three main results. First, the firm’s control over compatibility becomes irrelevant when the low-valuation market segment is sufficiently large. Second, when the size of the mass market and the quality improvement of the new good are both intermediate, the lack of commitment against future incompatibility *improves* the overall outcome by making it relatively more attractive for the firm to deviate from its short-term monopolistic behaviour, and to increase the size of the network by selling to low-valuation consumers. While a ban on incompatible upgrades increases profits in this case, welfare is higher if the monopolist is left to deal with its commitment problem on its own. Third and finally, when compatibility implies lower quality, the monopolist can be too reluctant to switch to a new, superior standard even when price-discrimination is possible and coordination failure is ruled out. In this last case the optimal policy intervention would be to *enforce*, not ban, incompatible upgrades.

To fix ideas, consider an unthreatened monopolist selling a software program whose users benefit from others using a compatible version.[2] Does Microsoft, for instance, have an incentive to make a new version of Word or Excel incompatible with its predecessor so that users are forced to upgrade only because everyone else does? It may be tempting, from one’s own personal experience, to answer in the affirmative. However, Microsoft has also been criticized for keeping its products compatible with an old standard whose “installed base” of existing users keeps the firm from switching to a new and superior (but incompatible) standard. In contrast, as Lerner and Tirole (2000) note, non-commercial ”open source” programmers, with no apparent incentive to induce obsolescence, seem to be more willing to abandon old standards. At the very least, this seems to suggest that arbitrary obsolescence is not the only possible outcome in a market where software is supplied by a (profit-seeking) monopolist.

This is not a case study of Microsoft. The aim of this paper is only to

[2]The unthreatened monopoly can be seen as a limiting case in a market where network externalities and proprietary standards create significant barriers to entry, or where standards are set by (implicit) cartels.

analyze some of the basic incentives for a monopolist with full discretion over intergenerational compatibility. This is done in a simple two-type setting where two generations of high-valuation consumers face a mass market of low-valuation users. As it turns out, this form of consumer heterogeneity is sufficient to overturn some qualitative results found in earlier work.

In the model, the existence of a low-valuation mass market leads to a trade-off for the monopolist between short-term profits and network size. Short-term profits are maximized by selling only to high-valuation consumers; in contrast, the size of the network (to which the firm implicitly sells access in later periods) is increased by selling the good to the low-valuation consumers in the mass market. This has implications for the firm's incentive to induce obsolescence by incompatibility, because the low-valuation consumers are less inclined than the high types to buy an upgrade to their existing goods; indeed, in the model, the low types' willingness to pay for an upgrade is zero. Hence, the firm cannot profitably shift the entire installed base to a new, incompatible standard, and leaving the low-valuation consumers behind reduces the premium which the high-types are willing to pay for an incompatible upgrade.

Heuristically, the results of this paper can be explained as follows. When the mass market is very large, its value to future high-type consumers dominates the short-term profit achievable from the early high types, so the firm sells to all consumers in period 1 and chooses compatibility later, deriving its profit mainly from selling access to a large network. The monopolist's discretion over compatibility is irrelevant in this case.

More interestingly, when the mass market is of intermediate size, the trade-off between short-term profits and network size is decisively affected by consumers' anticipation of inefficient, incompatible upgrades. If the firm cannot commit itself to compatibility, first-generation consumers reduce their willingness to pay for the original good, decreasing the opportunity cost for the monopolist to invest in network size by selling to the low-valuation mass market. Once it has done so, however, the incentive to choose incompatibility is reduced, leading to an overall outcome where the efficient output level in period one is followed by the efficient choice of compatibility in period two.

Finally, when compatibility implies lower product quality, the private incentive for the firm to choose compatibility can be too high from a welfare point of view. This is because under price discrimination, there are effictively two interdependent markets in the second period, one for the new good and one for the upgrade, and the optimal choice of compatibility and quantities is different for each. Profits in the market for the new good are maximized when all consumers adopt the new, superior standard, but this requires selling the upgrade to the low types owning the old version, reducing profits in the upgrade market. In a variation of the standard monopoly theme, the firm fails to maximize welfare because it targets the marginal buyer (in the upgrade market) rather than the average one.

This paper is related to a number of earlier articles. The analysis of the durable-goods monopolist competing with its own output in a market with a constant set of heterogeneous consumers goes back to Coase (1972) and Bulow (1982). Bulow (1986) describes planned obsolescence as a means for the monopolist to overcome this problem. Network externalities have been a popular topic since the seminal work of Katz and Shapiro (1985) and Farrell and Saloner (1985). For an overview, see e. g. Katz and Shapiro (1994); for a critical view, Liebowitz and Margolis (1994). Waldman (1993) and Choi (1994) describe planned obsolescence resulting from network effects and arbitrary incompatibility. The main difference between their work and the current paper is the introduction of heterogeneous consumers. Upgrades in a durable-goods monopoly without network effects are analyzed in Fudenberg and Tirole (1998). Ellison and Fudenberg (2000) find that banning upgrades can improve welfare in a model with network externalities and heterogeneous consumers. Waldman (1996) and Fishman and Rob (2000) consider innovation incentives in a durable-goods monopoly. Nahm (2000) analyzes forward and backward compatibility in a model with heterogeneous consumers and separate markets for hardware and software.

The model is introduced in the next section. Section 2.3 contains a benchmark case where the firm can credibly commit to compatibility. In Section 2.4, in contrast, this commitment is no longer possible. In Section 2.4.1 the mass market is shown to have a crucial effect on whether the firm's discretion over compatibility is relevant at all. Section 2.4.2 contains the case of intermediate mass market size and technical progress where the firm's lack of commitment actually improves the overall outcome from a welfare point of view. The possibility of inefficient compatibility is explored in Section 2.5 where the assumption of costless compatibility is given up. Section 2.6 concludes. Some proofs are in the Appendix.

## 2.2 The model

A monopolist sells a perfectly durable, indivisible good produced at zero cost to a set of non-atomic consumers. There is no threat of entry by other firms. The market exists for two periods $t = 1, 2$ of time; there is no discounting. In the first period, the good has intrinsic quality $\alpha_1$. Between periods, quality increases exogenously (and at no cost to the firm) to $\alpha_2 > \alpha_1$. There is no uncertainty, and all parameters are common knowledge. Before selling the improved good in period 2, the monopolist decides whether or not the two generations of goods will be compatible with each other; there is no fixed cost involved in this decision. Until Section 2.5 it will also be assumed that the intrinsic quality of the second-period good is independent of compatibility.

On the demand side, there are two types of consumers, high and low, whose per-period utility from using the good is a function of their type, the intrinsic quality of the good, and the number of consumers using a compatible good.[3] Thus, a consumer of type $\tau$ using a good of quality $\alpha$ for one period enjoys a utility of

$$u(\tau, \alpha, N)$$

if $N$ fellow consumers use a compatible product. When a consumer buys a new unit or an upgrade, he incurs a positive setup cost of $c$ before being able to use it.

For simplicity and to avoid a large number of less interesting subcases, the utility functions of the two consumers types are restricted as follows. Let the two types be denoted by $H$ (for high) and $L$ (for low), respectively. Then,

$$u(H, \alpha, N) = \theta\alpha + \beta N$$
$$u(L, \alpha, N) = c,$$

where $\alpha$ is the stand-alone (intrinsic) quality of the good, and $\beta N$ is the network value of a good when $N$ fellow consumers use a compatible product. The parameters $\theta$ and $\beta$ are both positive. Moreover, let

$$\theta\alpha_1 + \beta > c.$$

As a consequence, ($i$) it is profitable to provide the first-generation good to the high types for a single period; ($ii$) a social planner would have both types of

[3] As long as there is no risk of confusion, "number" and "mass" will be used interchangeably in this paper.

consumers use the good in both periods; but (*iii*) a profit-maximizing firm has no incentive to sell to low-type consumers in the second period because the maximimum price would be zero (so that after incurring the setup cost of $c$, the low types still have non-negative utility). For further reference,

$$\theta(\alpha_2 - \alpha_1) - c$$

will be called the *net quality increase* in period 2, i. e. the social value per high-type consumer of replacing the old good with the new.

In the first period, the mass of high-type consumers is normalized to unity. In addition, there are $\lambda$ low-type consumers. Let $H_1$ and $L_1$ denote, respectively, the sets of high- and low-type consumers entering the market at $t = 1$. Each consumer observes the price $p_1$ set by the monopolist for the first-period good and decides whether or not to buy one unit. Consumers have no use for a second unit of the good, and they cannot sell the good to others.[4] Since each individual's utility depends on the decisions taken by others, there may be multiple equilibria in the subgame where consumers choose to buy or abstain. As usual in the literature, consumers can coordinate on an outcome which is Pareto-undominated for them.

In the second period, the first-period consumers (and the goods which they have purchased) stay in the market. They are joined by a second generation of high and low types of the same size as the first, so that there is now a mass of 2 high-type consumers and $2\lambda$ low types. Let $L_2$ and $H_2$ denote the low and high types entering at $t = 2$. After the new consumers have arrived, the monopolist decides whether the second-period good will be compatible with the first. Compatibility is symmetric, i. e. a compatible good is both forward and backward compatible.[5] The firm then sets an upgrade price $p_2^U$ at which its former patrons can trade in their first-generation good for the new one, and a regular price $p_2$ at which the new good can be purchased by first-time buyers (i. e. second-generation customers and those from period 1 who did not buy the

[4]This is true, for instance, for most software markets.

[5]Ellison and Fudenberg (2000), among others, argue that it is more common for a newer program version to be forward incompatible but backward compatible; e. g. Word 2000 can read files created by Word 6.0 but not the other way around. On the other hand, even an owner of Word 2000 incurs a cost of forward incompatibility when he has to manage different file versions to exchange with others using an older program, so his utility should still increase in the number of fellow users with a fully compatible version. Besides, to the extent that the value of a network lies in the availability of complements – e. g. other users who can be asked for help – the distinction between forward and backward compatibility becomes less relevant. For a detailed analysis of forward and backward compatibility, see Nahm (2000).

old good). However, consumers cannot be forced to reveal their ownership of the old good, so the upgrade price can never exceed the price offered to first-time buyers.[6] Observing $p_2^U$ and $p_2$, consumers decide whether to upgrade (if they own the old good), buy the new good, or decline the firm's offer.

The solution concept is subgame-perfect Nash equilibrium.

## 2.3 Compatibility commitment

As a benchmark case, suppose the monopolist can credibly commit in period 1 to keeping the second-period product compatible. Solving the model backwards as usual, the monopolist sets the second-period prices $p_2^U$ and $p_2$ at $t = 2$ so as to maximize his second-period profit $\pi_2$ given the number of goods already in the market. Let $B_1 \subseteq \{L_1, H_1\}$ denote the installed base from period 1. With only two types of consumers, there are three cases at $t = 2$:

**Case 1:** $B_1 = \emptyset$ If the firm did not sell any unit of the first-generation good at $t = 1$, its only profitable option at $t = 2$ is to sell the new good to the high-type consumers $H_1$ and $H_2$, yielding a second-period (and total) profit of

$$\pi_2(\emptyset) = 2(\theta\alpha_2 + 2\beta - c). \tag{2.1}$$

**Case 2:** $B_1 = \{H_1\}$ Having sold the old product only to the high-type consumers $H_1$ in period 1, the firm sets a price $p_2 = \theta\alpha_2 + 2\beta - c$ for the newly-arrived high types $H_2$, extracting their entire rent. In addition, it will sell the second-generation good to its former customers as an upgrade if and only if

$$\theta(\alpha_2 - \alpha_1) - c > 0.$$

This is because the $H_1$ consumers facing an upgrade offer $p_2^U$ compare their utility from upgrading, $\theta\alpha_2 + 2\beta - c - p_2^U$, with their reservation utility from continued use of the old good, $\theta\alpha_1 + 2\beta$. Under compatibility, all the firm can charge for the upgrade is the perceived difference in stand-alone quality minus the setup cost which consumers incur when upgrading, so that

$$p_2^U = \theta(\alpha_2 - \alpha_1) - c \tag{2.2}$$

[6] This corresponds to the "semi-anonymous" case in Fudenberg and Tirole (1998).

This implies that a firm committed to compatibility takes the same upgrade decision as a social planner. The second-period profit is given by

$$\pi_2(H_1) = \theta\alpha_2 + 2\beta - c + \max\{\theta(\alpha_2 - \alpha_1) - c, 0\}. \tag{2.3}$$

**Case 3:** $B_1 = \{L_1, H_1\}$**.** Finally, if the firm has sold the old good to all consumers at $t = 1$, the second generation of high types $H_2$ is willing to pay up to $p_2 = \theta_2 + (2 + \lambda)\beta - c$, while the maximum upgrade price $p_2^U = \theta(\alpha_2 - \alpha_1) - c$ is the same as before. This gives a second-period profit of

$$\pi_2(L_1, H_1) = \theta\alpha_2 + (\lambda + 2)\beta - c + \max\{\theta(\alpha_2 - \alpha_1) - c, 0\}. \tag{2.4}$$

Turning to the first period, the firm maximizes the sum of its profits over both periods while consumers anticipate the outcome of the continuation game when they determine their willingness to pay for the first-period good.

Suppose the monopolist tries to sell the first-period good only to the high types $H_1$ present at $t = 1$. These consumers anticipate that they will replace the first good by an upgrade at $t = 2$ if $\theta(\alpha_2 - \alpha_1) - c > 0$, and that the second-generation high-types $H_2$ will join their network when they arrive. In equilibrium, $p_1$ must leave them indifferent between buying the first-period good and waiting until the second period (when their net utility will be zero because the monopolist fully extracts their rents). Thus,

$$p_1 = 2\theta\alpha_1 + 3\beta - c. \tag{2.5}$$

(Each consumer in $H_1$ has utility $\theta\alpha_1 + \beta - c$ in period 1 and $\theta\alpha_1 + 2\beta$ in period 2 if he does not upgrade; an upgrade increases his utility by $\theta(\alpha_2 - \alpha_1) - c$ but this additional rent is fully extracted by the upgrade price.) Note that the first-period good has a positive value in period 2 even if it is replaced, because it allows its owner to upgrade at a price reflecting only the incremental rather than the full value of the second-period good.

Given $p_1$, the total profit $\Pi(H_1)$ from selling only to the high-type consumers at $t = 1$ is

$$\begin{aligned}\Pi(H_1) &= p_1 + \pi_2(H_1)\\ &= 2\theta\alpha_1 + 3\beta - c + \theta\alpha_2 + 2\beta - c + \max\{\theta(\alpha_2 - \alpha_1) - c, 0\}\end{aligned} \tag{2.6}$$

In contrast, when the firm sells its first-period product to all $\lambda + 1$ consumers $L_1$ and $H_1$, it can achieve a higher second-period profit $\pi_2(L_1, H_1) > \pi_2(H_1)$ but has to set a price $p_1$ of no more than $c$. (The low types $L_1$ enjoy a gross utility of $2c$ over both periods but incur the setup cost of $c$ in period 1.) This yields a total profit from selling to all consumers in period 1 of

$$\Pi(L_1, H_1) = (\lambda + 1)c + \theta\alpha_2 + (\lambda + 2)\beta - c + \max\{\theta(\alpha_2 - \alpha_1) - c, 0\} \qquad (2.7)$$

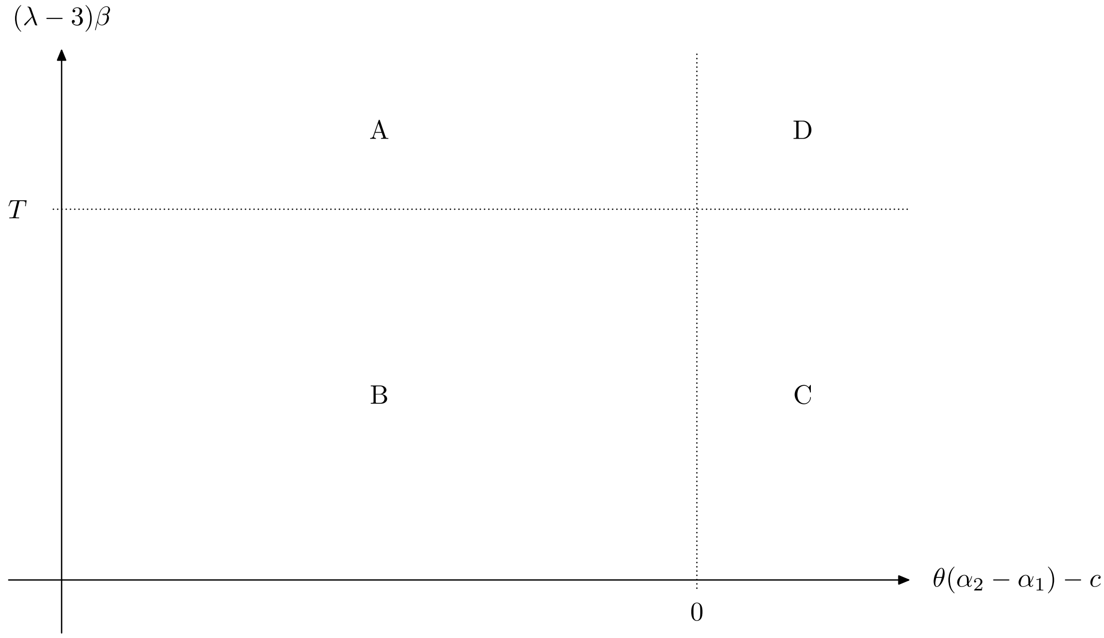


**Figure 2.1:** Overall outcomes with compatibility commitment. The firm sells to all first-period consumers if $(\lambda - 3)\beta$ is above the threshold $T \equiv 2\theta\alpha_1 - (\lambda + 2)c$ (areas A and D). An upgrade is sold to the first-generation high types $H_1$ if $\theta(\alpha_2 - \alpha_1) - c > 0$ (areas C and D).

Given the total payoffs from different levels of first-period output, the outcome under compatibility commitment is as follows.

**Lemma 1** *When the firm can commit (or is restricted) to compatibility, it sells the first-generation good only to the high-type consumers $H_1$ if*

$$(\lambda - 3)\beta < 2\theta\alpha_1 - (\lambda + 2)c$$

*and to all consumers $L_1$ and $H_1$ otherwise.*

**Proof.** When the firm does not sell the first-period good to any consumers, its total profit is given by (2.1). This is strictly less than $\Pi(H_1)$ from (2.6). The

Lemma follows from the comparison of (2.6) with (2.7). ■

Intuitively, the firm faces a trade-off between the rents which it extracts from its first-period consumers and the size of the network to which it implicitly sells access in period 2. Restricting its first-period output to the high-valuation types $H_1$ increases its first-period profit; this is the standard monopoly effect. In contrast, the value of the network is larger for the second-period high types $H_2$ if it includes the low-valuation "mass market" $L_1$.[7] This network effect dominates the monopoly effect if the network value of the first-period low types $L_1$ to the second-generation high types $H_2$, $\lambda\beta$, is large relative to the rent which can be extracted form the first-generation high types, $2\theta\alpha_1 + 3\beta - c$, when first-period output is restricted to them. This is depicted by areas A and D in Figure 2.1, where the term

$$(\lambda - 3)\beta$$

measures the difference between the network contribution of the low types $L_1$ and the opportunity cost of giving the high types $H_1$ free access to the network.

Compatible upgrades are efficient because the upgrade price $p_2^U$ in (2.2) is equal to the net incremental value of the upgrade relative to the perfectly durable (and compatible) old good.

**Lemma 2** *A compatible upgrade is sold to the high-type consumers $H_1$ at $t = 2$ if and only if upgrading is efficient, i. e. iff*

$$\theta(\alpha_2 - \alpha_1) - c > 0.$$

## 2.4 Discretionary incompatibility

In contrast with the benchmark case in the previous section, suppose the monopolist can choose at $t = 2$ whether or not the new product will be compatible with the existing installed base. As Choi (1994) shows, this allows the firm to set a higher upgrade price than under compatibility when consumers are homogeneous (so that they all choose to upgrade together). In particular, by reducing the network value of the old good, incompatibility can make upgrades profitable to sell when they are socially wasteful, i. e. when $\theta(\alpha_2 - \alpha_1) - c < 0$.

[7] Delaying all sales until the second period is always dominated because the firm can fully extract the high-type consumers' rents when selling to them from period 1; this is a consequence of the assumption that the firm can discriminate between its former patrons and new customers. Cf. Choi (1994).

The purpose of this section is to show that $(i)$ the incompatibility option is not used in equilibrium when the network value of the low-type consumers (to the high types) lies above a threshold, and $(ii)$ the lack of commitment against incompatibility has the beneficial effect of lowering this threshold precisely when incompatible upgrades are inefficient. As in the previous section, the behaviour of the firm in period 2 depends on whom it sold the first-period product.

**Case 1:** $B_1 = \{\emptyset\}$**.** If the firm has not sold any units of the first-period product, compatibility is irrelevant. The monopolist sells the second-period good to both high-type groups $H_1$ and $H_2$ and achieves the same second-period (and total) profit $\pi_2(\emptyset)$ from (2.1) as before.

**Case 2:** $B_1 = \{H_1\}$**.** After selling only to the high types at $t = 1$, the monopolist can keep the new good compatible and achieve the second-period profit $\pi_2(H_1)$ from (2.3). Alternatively, it can choose incompatibility to raise the upgrade price which its former patrons are willing to pay. Comparing their utility from upgrading, $\theta\alpha_2 + 2\beta - c$, with their reservation utility of using the old, incompatible product, $\theta\alpha_1 + \beta$, the first-generation high types $H_1$ are indifferent when

$$p_2^{U,IC} = \theta(\alpha_2 - \alpha_1) + \beta - c, \tag{2.8}$$

where the superscript $IC$ indicates incompatibility. (Note that compared with $p_2^U$ in (2.2), this upgrade price is increased by $\beta$, reflecting the decrease in utility of the old good due to incompatibility.) Thus, the firm can make a positive profit from selling incompatible upgrades whenever

$$\theta(\alpha_2 - \alpha_1) - c > -\beta.$$

As long as this is the case, the newly-arrived $H_2$ users pay up to $\theta\alpha_2 + 2\beta - c$ for the second-period good, the same as under compatibility. This leads to a second-period profit of

$$\pi_2^{IC}(H_1) = \theta\alpha_2 + 2\beta - c + \theta(\alpha_2 - \alpha_1) + \beta - c.$$

**Lemma 3** *When incompatibility is possible and the first-period good has been sold only to the high types $H_1$, the second-period outcome is*

$$\begin{cases} \textit{Compatibility and no upgrades} & \textit{if} \quad \theta(\alpha_2 - \alpha_1) - c \in (-\infty, -\beta) \\ \textit{Incompatibility and inefficient upgrades} & \textit{if} \quad \theta(\alpha_2 - \alpha_1) - c \in [-\beta, 0), \\ \textit{Incompatbility and efficient upgrades} & \textit{if} \quad \theta(\alpha_2 - \alpha_1) - c \in [0, \infty). \end{cases}$$

*The new high-type consumers $H_2$ always buy the second-period good at*

$$p_2 = \theta\alpha_2 + 2\beta - c.$$

**Proof.** When $\theta(\alpha_2 - \alpha_1) - c < -\beta$, the firm cannot profitably sell even an incompatible upgrade to its earlier customers, and incompatibility would reduce the first-time buyers' willingness to pay. Therefore, the monopolist keeps the new good compatible and sells it only to the new high types $H_2$.

When $\theta(\alpha_2 - \alpha_1) - c \in [-\beta, 0)$, an incompatible upgrade is profitable to sell ($p_2^{U,IC} > 0$) even though the perceived increase in quality, $\theta(\alpha_2 - \alpha_1)$, is less than the setup cost incurred by those who upgrade, so the outcome is inefficient. (A compatible upgrade selling at $p_2^U = \theta(\alpha_2 - \alpha_1) - c$ would not be profitable.)

When $\theta(\alpha_2 - \alpha_1) - c > 0$, incompatibility still raises the upgrade price by $\beta$ (so the firm prefers incompatibility), but there is no efficiency loss.

Regardless of the upgrade decision, first-time customers are sold the second-period product plus access to the network consisting of both high-type groups $H_1$ and $H_2$, so $p_2$ is invariant. ■

**Case 3:** $B_1 = \{L_1, H_1\}$**.** When the installed base comprises all first-generation consumers, incompatibility is relatively less profitable at $t = 2$ because the low-type consumers $L_1$ will not upgrade at any positive price (since their valuation of both goods is the same), and leaving them behind reduces the utility of the high-type consumers buying the new, incompatible product.

**Lemma 4** *Suppose incompatibility is possible and $\lambda > 1/2$. If the first-period good has been sold to both $L_1$ and $H_1$, the second-period outcome is compatibility and an efficient upgrade decision. The firm achieves the second-period profit of $\pi_2(L_1, H_1)$ from (2.4).*

**Proof.** Under compatibility, the second-period profit is given by (2.4). Under incompatibility, the new high-type consumers $H_2$ pay at most $\theta\alpha_2 + 2\beta - c$ (i. e. $\lambda\beta$ less than under compatibility). For the first-generation high types $H_1$, the upgrade is worthwhile up to a price of $p_2^U = \theta(\alpha_2 - \alpha_1) + (1 - \lambda)\beta - c$ (this follows from comparing their utility derived from the new product in a network of $H_1$ and $H_2$ with their reservation utility from the old good in a network of $L_1$ and $H_1$.) Hence,

$$\pi_2^{IC}(L_1, H_1) = \theta\alpha_2 + 2\beta - c + \theta(\alpha_2 - \alpha_1) + (\lambda + 1)\beta - c,$$

which is less than $\pi_2(L_1, H_1)$ from (2.4) when $\lambda > 1/2$. ■

In what follows, it will be assumed that $\lambda$ is indeed larger than 1/2; this is not a very restrictive assumption given the focus of this paper on a large low-valuation market segment.

Given the continuation payoffs from $t = 2$, the firm again chooses its first-period output (via the price $p_1$ which it sets) in order to maximize its total profit $\Pi$ over both periods, as will be discussed in the following two subsections.

### 2.4.1 Obsolescence and the mass market

If the firm sells the first-generation good only to the high types $H_1$, consumers expect (by Lemma 3) that it will introduce an incompatible upgrade in period 2 whenever $\theta(\alpha_2 - \alpha_1) - c > -\beta$. They also anticipate that incompatibility will increase the utility difference between the new good and the old by $\beta$, and thus raise the upgrade price by the same amount (cf. (2.2) and (2.8)). This reduces the value of the first-period good as an upgrade base in period 2, so when incompatibility is expected the first-period high types pay at most

$$2\theta\alpha_1 + 2\beta - c$$

for the first-period good (i. e. $\beta$ less than under a commitment to future compatibility, cf. (2.5)). The resulting overall profit is

$$\Pi(H_1) = \begin{cases} 2\theta\alpha_1 + 3\beta - c + \theta\alpha_2 + 2\beta - c & \text{if } \theta(\alpha_2 - \alpha_1) - c < -\beta, \\ \begin{array}{l} 2\theta\alpha_1 + 2\beta - c + \theta\alpha_2 + 2\beta - c \\ \quad + \theta(\alpha_2 - \alpha_1) + \beta - c \end{array} & \text{otherwise.} \end{cases} \tag{2.9}$$

Alternatively, the firm can decide to sell the first-generation product to all consumers $L_1$ and $H_1$. By Lemma 4 and the assumption that $\lambda > 1/2$, the final outcome is the same as if the firm was committed to compatibility, with the overall profit $\Pi(L_1, H_1)$ given by (2.7). The comparison of (2.7) with (2.9) yields the following result.

**Proposition 1** *Suppose the firm is free to make the second-period good incompatible with the one sold before. If*

$$(\lambda - 3)\beta > 2\theta\alpha_1 - (\lambda + 2)c \tag{2.10}$$

*it sells the first-period good to all consumers $L_1$ and $H_1$ and never chooses incompatibility at $t = 2$. The inequality in (2.10) is sufficient but not necessary.*

The intuition is essentially the same as when the firm is committed to second-period compatibility. The first-period trade-off for the firm is between extracting the rents of the high types $H_1$ (by keeping prices high and output low), and maximizing the size of the network (by selling to all first-period consumers). After selling only to the high types in period 1, incompatibility increases the second-period upgrade price. However, this is anticipated in period 1, so the overall profit of the low-output strategy is no higher than under compatibility. Thus, when the network value of the "mass market" low types $L_1$ to the second-period high types is large, the network effect again dominates the monopoly effect (areas A and D in Figure 2.1), turning the incompatibility option irrelevant. It should be noted, however, that this depends on the existence of a large set of low-valuation consumers whose willingness to pay for an upgrade (even an incompatible one) is too low for the monopolist to profitably shift the entire market to the new product in period 2.

### 2.4.2 Lack of commitment and efficiency

Proposition 1 states that the lack of a commitment against discretionary incompatibiliy is irrelevant when the size of the low-valuation market segment lies above a threshold value; this corresponds to areas A and D in Figures 2.1 and 2.2. In addition, it turns out that the commitment problem has a beneficial effect, from a planner's point of view, by lowering this threshold precisely when discretionary incompatibility leads to inefficient upgrades. This will be the focus of the present subsection.

The key idea is that the anticipation of inefficient upgrades has an impact on the first-period trade-off between extracting consumers' rents and maximizing the size of the network. By Lemma 3, consumers expect inefficient upgrades when ($i$) the first-period good has been sold only to the high types $H_1$, and ($ii$) the net quality increase $\theta(\alpha_2 - \alpha_1) - c$ is negative but larger than $-\beta$, so that the price $p_2^{U,IC}$ of an incompatible upgrade in (2.8) is still positive. Since the firm extracts the entire rent of the high types $H_1$ when it sells the first-period good only to them, it also bears the full welfare loss caused by inefficient upgrades. Thus, the low-output strategy becomes relatively less profitable in period 1 while the second-period value of a larger network including the low types $L_1$ is unchanged. Consequently, the incentive for the monopolist to sell to all consumers in period 1 (and to choose compatibility at $t = 2$) is increased whenever incompatibility leads

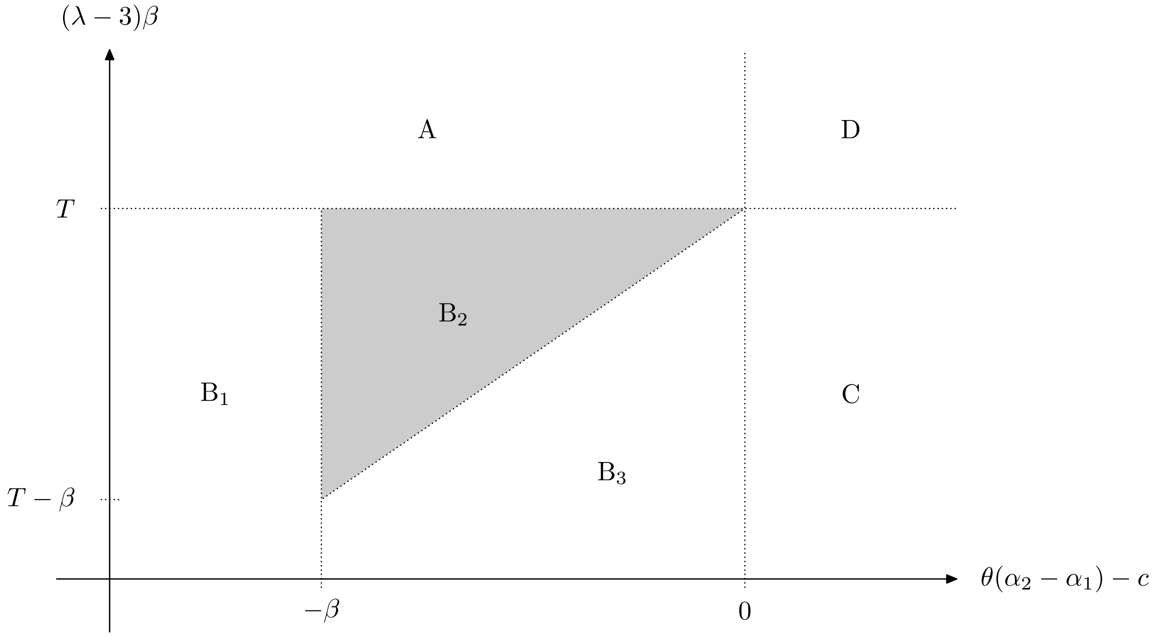


**Figure 2.2:** Overall outcomes when the firm has discretion over incompatibility. $T \equiv 2\theta\alpha_1 - (\lambda + 2)c$ denotes the threshold value for $(\lambda - 3)\beta$ from Figure 2.1. The shaded area $B_2$ represents the case where the commitment problem leads to higher first-period output. In area $B_3$, incompatibility leads to inefficient upgrades. Outside of $B_2$ and $B_3$ the outcome is the same as in the commitment case.

to inefficient upgrades.

Suppose $(\lambda - 3)\beta$ lies below the threshold $T$ in Figure 2.2, so that Proposition 1 does not apply. When the net quality increase $\theta(\alpha_2 - \alpha_1) - c$ is below $-\beta$ (area $B_1$), the outcome is the same as in the benchmark commitment case. The first-period good is sold only to the high types $H_1$ because the monopoly effect dominates the network effect; there are no upgrades since the incremental value of even an incompatible second-period good is below the setup cost $c$; and the second-period high types are offered a compatible good when they arrive on the market. Similarly, when the net quality increase is positive (area C), the firm sells the first good only to the high types $H_1$ (again, the monopoly effect dominates the network effect at $t = 1$), chooses incompatibility at $t = 2$ and sells the new product to both high-valuation groups. The first-period customers are "forced" to buy an incompatible upgrade but there is no loss of efficiency; nor are consumers any worse off, since the higher upgrade price under incompatibility is compensated by a lower first-period price. Thus, when the exogenous quality increase measured by $\theta(\alpha_2 - \alpha_1)$ is either very small or very large, the firm's discretion over compatibility has no welfare effects.

The interesting situation is the one where the net quality increase $\theta(\alpha_2 - \alpha_1) - c$ is negative but the firm can nevertheless charge a positive upgrade price for an incompatible second-period product. In Figure 2.2, this case is represented by areas $B_2$ and $B_3$.

By Lemmas 3 and 4, the second-period outcome is incompatibility and inefficient upgrades if first-period sales have been restricted to the high types $H_1$, and compatibility and no upgrades if the low types $L_1$ also bought the first-period good. Given this, the profit-maximizing output level in period 1 again depends on whether the relative network value of the low types, $(\lambda - 3)\beta$, lies above a threshold value. However, this threshold is lower than the corresponding critical value in the commitment case (represented by the line $T$ in Figure 2.1).

**Proposition 2** *Let $\theta(\alpha_2 - \alpha_1) - c \in [-\beta, 0)$, and let $T \equiv 2\theta\alpha_1 - (\lambda + 2)c$. When the firm is free to choose incompatibility, it*

- *sells to all consumers $L_1$ and $H_1$ in period 1 and*
- *chooses compatibility in period 2*

*if and only if $(\lambda - 3)\beta$ lies above a threshold value given by*

$$T + \theta(\alpha_2 - \alpha_1) - c.$$

**Proof.** Comparing (2.7) with (2.9), the firm prefers selling to all first-period consumers if and only if

$$(\lambda - 3)\beta > \theta\alpha_1 + \theta\alpha_2 - (\lambda + 3)c.$$

The RHS is equal to $T + \theta(\alpha_2 - \alpha_1) - c$. ■

Thus, compared with the commitment case in Figure 2.1, the firm's discretion over compatibility leads to two very different new outcomes when the relative size of the mass market and technological progress are both intermediate. On the one hand, when the relative network value of the low types $L_1$ is small, arbitrary obsolescence reduces both welfare and profits; this is the result found by Choi (1994) and represented by area $B_3$ in Figure 2.2. On the other hand, the trade-off between the monopoly effect and the network effect is shifted towards the latter, lowering the threshold value for $(\lambda - 3)\beta$ above which the monopolist chooses the efficient output level (i. e. sells to all consumers) in period 1. This creates area $B_2$ in Figure 2.2, where sales to all consumers $L_1$ and $H_1$ in the first period are followed by second-period compatibility and the (efficient) decision not to offer upgrades. The border between $B_2$ and $B_3$ represents the new, lower threshold value of $T + \theta(\alpha_2 - \alpha_1) - c$.

From a theoretical perspective, Proposition 2 underscores the role of consumer heterogeneity in this model of network externalities and compatibility. The beneficial effect of the commitment problem in area $B_2$ arises because the anticipation of incompatibility reduces the first-period willingness to pay of the "marginal" consumer in $H_1$, weakening the standard monopoly effect relative to the (unchanged) network effect. With homogeneous consumers, this cannot occur.

From a policy perspective, Proposition 2 implies that consumer heterogeneity makes the welfare effect of a ban on incompatibility ambiguous. In the situation represented by area $B_2$, welfare is higher when the firm is left to handle its commitment problem on its own; mandatory compatibility increases profits but hurts consumers.[8] On the other hand, banning incompatibility increases both profits and welfare in the case depicted by area $B_3$.

## 2.5 Inefficient compatibility

Up to this point, there was no reason for a social planner to prefer incompatibility since it implied neither lower costs nor higher quality of the second-period good. In this section, in contrast, compatibility is assumed to imply lower quality, as it often does in an industry with technological progress. Suppose that the monopolist can choose between two different goods in period 2, one of which is compatible with the old product while the other is not. Let $\alpha_2^C$ denote the inherent quality of the compatible second-period good, and let $\alpha_2^{IC}$ stand for the level of quality which can be attained without being constrained to compatibility. By assumption,

$$\alpha_2^{IC} > \alpha_2^C > \alpha_1.$$

[8] Note that by increasing its output at $t = 1$, the firm does not *solve* its commitment problem. Only the first-period high types are affected by the lack of commitment; increasing sales in period 1 *reduces* the rent extracted from these consumers. Creating a larger installed base is not a commitment device but an investment which becomes relatively more attractive when the threat of incompatibility is relevant.

Thus, while both second-period goods are better than the old one, compatibility reduces the inherent quality of the new product.

The main result in this section is that there can be too much compatibility, and that forcing the firm to choose the incompatible product can increase welfare. In the second period the firm is active in two interdependent sub-markets, upgrades and new goods, and the profit-maximizing choice of compatibility and quantities is different for each. Profits in the market for new goods are maximized when all consumers adopt the new standard, but this requires selling the upgrade to the low types $L_1$, reducing profits in the upgrade market. The possible loss of overall welfare stems from the well-known fact that the monopolist targets the marginal buyer (in the upgrade market) whereas a planner would maximize the utility of the average consumer.

The following is not a complete description of every possible constellation but a demonstration of the welfare benefits of mandatory incompatibility in a reasonably general example. Thus, the attention will be restricted to the case characterized by the following two assumptions.

**Assumption 1**

$$\theta(\alpha_2^C - \alpha_1) - c > \theta(\alpha_2^{IC} - \alpha_2^C).$$

This implies that (*i*) both types of upgrades are viable in the sense that the incremental quality exceeds the setup cost, and (*ii*) the compatible upgrade embodies most – but not all – of the exogenous technological progress which has occurred between periods.

**Assumption 2**

$$(\lambda - 3)\beta > \theta\alpha_1 + \theta\alpha_2^{IC} - 2c,$$

so that the value of the first-period low types $L_1$ to the second-period high types $H_2$, given by $\lambda\beta$, is large relative to the rent which can be extracted from the first-generation high-types $H_1$ by selling only to them in period 1 and offering the incompatible upgrade in the second period.

Under these assumptions, an unconstrained monopolist chooses compatibility in period 2.

**Lemma 5** *Under Assumptions 1 and 2, the monopolist*

- *sells the first-period good to all consumers* $L_1$ *and* $H_1$ *present at* $t = 1$*,*
- *chooses the compatible good at* $t = 2$*,*

- *and sells the new good to the high-type consumers $H_1$ (as an upgrade) and $H_2$ (as a full product).*

A detailed proof is given in the Appendix. Essentially, by Assumption 2, it pays for the monopolist to forgo extracting the rent of the first-period high types, and to invest in network size instead by selling to both low- and high-valuation consumers $L_1$ and $H_1$ in period 1. In period 2, the firm can choose to maximize the total quality of the good which it sells to the newly-arrived high types $H_2$ by selling the incompatible upgrade to all period-1 consumers; however, this implies that the marginal consumer in the upgrade market has a valuation of zero for the new good. Alternatively, the firm can increase its profit in the upgrade market by selling only to high-valuation consumers, but it has to choose the lower-quality compatible good in this case in order to preserve the installed base it created in period 1. Under Assumptions 1 and 2, it is indeed more profitable to sell fewer upgrades of lower quality at a positive price than to give away a better, incompatible upgrade to all former customers, even though this implies selling a lower-quality product to the new consumers. This is a variation of the well-known monopoly theme of the firm taking into account the marginal consumer rather than the average one, so it is not surprising that the outcome can be socially inefficient.

Now suppose the firm is forced to choose the incompatible, higher-quality product in period 2. As the following Lemma states, the outcome is the same as before except that the upgrade is sold to the low-type consumers $L_1$ as well.

**Lemma 6** *Under Assumptions 1 and 2, if the firm is constrained to incompatibility, it*

- *sells the first-period good to all consumers $L_1$ and $H_1$ present at $t = 1$,*
- *"sells" the incompatible upgrade to all first-period consumers $L_1$ and $H_1$ at a price of zero, and*
- *sells the new product to the second-period high types $H_2$.*

A detailed proof, similar to the one for Lemma 5, is given in the Appendix. Intuitively, Assumption 2 guarantees that in period 1 the monopolist prefers to maximize the size of the network (to which is sells access in the second period) even if this implies having to provide free upgrades in period 2.

It is now straightforward to derive the welfare effect of mandatory incompatibility.

**Proposition 3** *Under Assumptions 1 and 2, the welfare effect of restricting the monopolist to incompatibility is given by*

$$\Delta W = 2\theta(\alpha_2^{IC} - \alpha_2^{C}) - \lambda c.$$

**Proof.** Since the cost of production is zero, welfare is equal to the total utility enjoyed by each consumer group over both periods. The Proposition follows directly from Lemmas 5 and 6. ■

Intuitively, the welfare effect of mandatory incompatibility is positive when the incremental quality increase enjoyed by the high types $H_1$ and $H_2$ is larger than the additional setup cost incurred by the low-valuation consumers $L_1$. Thus, in this simple model of network externalities and heterogeneous consumers there can be "excess inertia" (a tendency for an old standard to prevail when it should be replaced) even though by assumption there is no coordination failure, the standard is "sponsored" (owned by the monopolist), and the firm can price discriminate (albeit imperfectly).

## 2.6 Conclusion

This paper has highlighted the role of consumer heterogeneity in a durable-goods monopoly where the firm has full control over compatibility between successive product generations. In particular, the focus has been on the effect of a "mass market" of low-valuation consumers existing alongside a set of high-valuation users. When consumers are heterogeneous, the standard monopoly effect comes into play, interacting with the network effect and affecting the qualitative outcome in a non-trivial way.

There are three results. First, when the mass market is large (and comprised of consumers whose valuation for quality is sufficiently low), the firm's discretion over compatibility is irrelevant. The firm forgoes extracting the rents of its early high-valuation customers, expanding output instead to increase the size of the network to which it can implicitly sell access later on.

Second, when the size of the mass market and the speed of technological progress are both intermediate, the lack of commitment against ex-post incompatibility increases the incentive to invest in network size early on, leading to a more efficient overall outcome precisely when there would be inefficient upgrades in a homogeneous-consumers world. This is yet another instance of one inefficiency (the potential arbitrary obsolescence of the old good) counteracting another (the standard monopoly effect).

Finally, consumer heterogeneity can lead to the firm choosing compatibility when it would be efficient, from a welfare point of view, to switch to a superior but incompatible new standard.

As they stand, these results have clear policy implications. In earlier research, a ban on incompatible upgrades could increase welfare. The results in this paper suggest that when there is a large mass market, (*i*) the problem of discretionary incompatibility may be irrelevant; (*ii*) when it is not, it can be more efficient to let the firm act on its commitment problem by expanding output than to regulate its compatibility choice; and (*iii*) when compatiblity implies lower quality, under some conditions the best policy is to make incompatible upgrades mandatory.

However, these results should not be read as immediate policy recommendations. The model employed in this paper is quite simplistic, and a number of parameter assumptions have been made in order to concentrate on the more interesting cases. Given the strong qualitative effects found so far, it should be worth further to explore the role of consumer heterogeneity in a more general setting with a richer set of consumer types. In particular, it would be interesting to see the implications for R&D incentives in an infinite-horizon framework like in Fishman and Rob (2000). This is left for further research.

## 2.7 Appendix

**Proof of Lemma 5**

The first step of the proof is to derive the optimal compatibility and output choice of the firm at $t = 2$. The second step is to let the firm maximize its total profit at $t = 1$ given its subgame-perfect behaviour from step 1.

At $t = 2$, given an installed base of $L_1$ and $H_1$, the firm can

1. choose compatibility and sell to $H_1$ and $H_2$ for a second-period profit of

$$\theta(\alpha_2^C - \alpha_1) - c + \theta\alpha_2^C + (\lambda + 2)\beta - c;$$

2. choose incompatibility and sell to $H_1$ and $H_2$ for a second-period profit of

$$\theta(\alpha_2^{IC} - \alpha_1) + (1 - \lambda)\beta - c + \theta_2^{IC} + 2\beta - c, \tag{2.11}$$

or

3. choose incompatibility and sell to $L_1$, $H_1$, and $H_2$ for a second-period profit of

$$\theta\alpha_2^{IC} + (\lambda+2)\beta - c. \tag{2.12}$$

(The low-type consumers $L_1$ and $L_2$ pay at most zero for the new good, so all other options are dominated.) The first option domintates the second if

$$(2\lambda-1)\beta > 2\theta(\alpha_2^{IC} - \alpha_2^{C}),$$

which is implied by Assumption 2. By Assumption 1, the first option also dominates the third. The overall profit over both periods from selling to $L_1$ and $H_1$ at $t = 1$ is therefore (cf. (2.7))

$$\Pi(L_1, H_1) = (\lambda+1)c + \theta\alpha_2^{C} + (\lambda+2)\beta - c + \theta(\alpha_2^{C} - \alpha_1) - c.$$

At $t = 2$, given an installed base of only $H_1$, the optimal choice is incompatibility and sales of the second-period good to $H_1$ and $H_2$ (Lemma 3), resulting in an overall profit over both periods of (cf. (2.9))

$$\Pi(H_1) = 2\theta\alpha_1 + 2\beta - c + \theta\alpha_2^{IC} + 2\beta - c + \theta(\alpha_2^{IC} - \alpha_1) + \beta - c. \tag{2.13}$$

Finally, if the firm sells nothing at $t = 1$ its overall profit is

$$\Pi(\emptyset) = 2(\theta\alpha_2^{IC} + 2\beta - c).$$

Maximizing its overall profit at $t = 1$, the firm never chooses to sell nothing (which is always dominated by selling to $H_1$), and prefers to sell to all consumers if

$$(\lambda-3)\beta > 2\theta(\alpha_2^{IC} - \alpha_2^{C}) + 2\theta\alpha_1 - (\lambda+2)c,$$

which holds by Assumptions 1 and 2. ■

**Proof of Lemma 6**

The incompatibility constraint affects the firm only when has sold the first-period good to all consumers $L_1$ and $H_1$ present at $t = 1$. In this case, the only relevant choice under incompatibility is whether to sell the upgrade to the low types $L_1$ as well as the high types $H_2$. By (2.11) and (2.12), the firm sells the upgrade to both groups if

$$(2\lambda-1)\beta > \theta\alpha_2^{IC} - \theta\alpha_1 - c,$$

which is assured by Assumption 2. This gives an overall profit over both periods of selling to $L_1$ and $H_1$ at $t = 1$ of

$$\Pi(L_1, H_1) = \theta\alpha_2^{IC} + (\lambda + 2)\beta - c \tag{2.14}$$

(note that the low types $L_1$ have a valuation (net of the setup cost) of zero for one-period consumption, so both the first-period good and the upgrade must be sold at a price of zero.) Comparing (2.14) with the overall profit (2.13) from selling the first-period good only to $H_1$, the firm prefers to sell to all first-period consumers if and only if Assumption 2 holds. ■

# Chapter 3

# Interconnection, Termination-Based Price Discrimination, and Network Competition in a Mature Telecommunications Market

## 3.1 Introduction

A distinctive feature of the telecommunications industry is the need for competing networks to interconnect in order to provide each of their customers with access to the entire consumer population. This leads to an interesting coexistence of vertical and horizontal relationships, where each network is the monopolistic supplier of an intermediate input—the termination of calls directed towards its own subscribers—to its rivals in the retail market, and in turn relies on them to provide access to their customers. It is generally agreed that the outcome will be socially inefficient if each firm sets its termination fee independently, because a unilateral increase by any firm raises its rivals' costs without affecting its own. Thus, unless the fee for terminating calls is set by a regulator, efficiency demands that the firms be allowed to cooperate in setting a common, reciprocal price for this mutually-provided intermediate input. This immediately raises the question of whether an unregulated access charge can be an instrument of collusion in the retail market.

It is fairly clear that interconnection conditions need to be regulated as long as there is one dominant player in the market. While collusion in the retail market is not usually a major issue in this situation, the incumbent has an obvious incentive to erect entry barriers by raising interconnection prices. However, there is a widespread perception that this sector-specific regulation is necessary and justified only until the market is "mature" in the sense that no single player any longer holds a dominant position. Legislation in Germany, for instance, provides for sector-specific regulation to be replaced by regular antitrust provisions once the market is in a state of "workable competition". Unfortunately, it is not obvious from a theoretical point of view that such competition, however defined, in the retail market will in itself assure an efficient outcome when the firms are allowed to set the terms and conditions of interconnection cooperatively.

The main point of the present paper is to show how control over the access charge facilitates collusion in the retail market even when the firms compete in non-linear tariffs, as long as they can price-discriminate according to where calls terminate. This is in contrast with a common observation in the literature that non-linear tariffs eliminate the collusive scope of the access charge. Moreover, when consumers' calling patterns are sufficiently biased, the model predicts a *markup* on the price of calls to other networks—a typical characteristic of real-life telecommunications markets which most existing models on non-linear tariffs fail to explain. As for policy implications, the results of this paper suggest that standard antitrust provisions alone will not necessarily ensure efficiency even in mature telecommunications markets. On the other hand, sector-specific regulation might achieve its goal of preventing collusion simply by banning termination-based price discrimination, rather than by setting interconnection fees outright. Finally, on the theoretical level, this paper demonstrates how the degree of horizontal differentiation between networks can be the endogenous outcome of the firms' strategic interaction, based on a plausible assumption about consumer behaviour.

The standard model of interconnection in a mature market is described in several variations by Laffont et al. (1998a) and Laffont et al. (1998b). Other papers on the subject include Armstrong (1996), Armstrong (1998) and Carter and Wright (1999), employing a very similar approach. Briefly stated, these are static, discrete-choice models of the Hotelling type, where a set of atomless consumers is distributed over the characteristics space between two firms, each consumer is connected to the entire network by one of the firms, and consumers choose their calling partners with equal probability among the entire consumer population. In this kind of setup, the firms' ability to collude by manipulating the reciprocal access charge depends on the type of tariffs offered to the final customers. In particular, a simple linear retail tariff results in a common incentive to raise the access charge above marginal cost, which raises the cost of cutting prices to win market share, and thus enables the networks to sustain a more profitable shared-market equilibrium. On the other hand, the scope for collusion disappears with non-linear tariffs, which allow each firm to attract additional consumers by reducing their fixed fee without incurring an access deficit from lowering usage prices. Equilibrium profits are then only a function of the exogenous "transportation" cost in the Hotelling setup, leaving the networks indifferent about the interconnection fee. This result continues to hold when consumers differ in their preferred call volumes, as in Dessein (2000) and Hahn (2000). Since two-part and other non-linear tariffs are pervasive in telecommunications, this seems to support the view that sector-specific regulation might be only a temporary necessity.

However, real-life tariffs are often not just non-linear, but they also discriminate between calls to members of the same network ("on-net" calls) and calls terminating on another network ("off-net" or "outbound" calls). This is particularly true in most mobile-telephony markets. Laffont et al. (1998b) analyze this case of termination-based price discrimination. As shown by Gans and King (1999), the result is a cooperative agreement to sell access at a *discount*. (Among other things, this implies that "bill and keep" arrangements can be a form of imperfect but simple collusion.) The argument goes as follows. With access to the other network sold below marginal cost, firms will offer outbound calls to their consumers at a lower rate than on-net calls. Other things equal, this means consumers prefer subscribing to a smaller network. This in turn makes it less attractive for each firm to expand its consumer base, since the marginal consumer has to be compensated for the increase in the size of the network he subscribes to. Therefore, the firms can sustain a profitable shared-market equilibrium with an access discount and off-net rates below on-net prices.

This argument, however, is somewhat at odds with casual empiricism. In many mobile-telephony markets, for instance, calls to other networks are substantially more expensive than on-net calls. To the extent that interconnection fees and costs are publicly observable, it also seems highly doubtful that most networks terminate each others' outbound calls at a discount. Moreover, from a theoretical perspective, it is not obvious that the assumption of exogenous horizontal differentiation is quite appropriate in a market like telecommunications where the basic good—the transmission of signals—is essentially a homogeneous commodity. One might argue that the good which consumers really demand is network membership, and that the networks differ in the features and services they offer. However, these are endogenous choices for the networks, and it is not clear that they would choose to place themselves at their respective end of the Hotelling line; anecdotal evidence suggests on the contrary that networks tend to match their rivals' features quite closely.

This paper introduces two innovations to the basic framework of the standard model. First, consumers' preferences for the two networks are no longer exogenously given, but a function of calling patterns and tariffs. Thus, the perceived horizontal differentiation upon which the profitability of a shared-market outcome depends becomes a result of the firms' strategic behaviour rather than an a-priori assumption. Second, it is assumed that consumers do not choose their calling partners randomly among all others, but that they make most of their calls to a specific subset of consumers (their "group"). The main result of the paper is that the access charge can be used as a collusive device even when the networks compete in non-linear tariffs as long as termination-based price discrimination is possible. In particular, an interconnection *markup* (and the resulting markup on off-net calls) can be used to differentiate the otherwise homogeneous networks if consumers' bias towards their preferred calling partners is sufficiently strong.

Heuristically, the main effect can be explained in a very simple model of spatial competition. Suppose there are two beaches connected by a road, and an ice-cream vendor and a set of consumers on each of them. Everything is symmetric. The vendors can neither move their location nor discriminate among consumers. There is a transportation cost of $t$ for each consumer walking to the other beach to buy his ice-cream there. Consumers have unit demand and a reservation price of $r$. It is well-known that in this setting the only pure-strategy Nash equilibrium (*i*) is symmetric; (*ii*) exists if and only if $t > (r - \phi)/2$, where $\phi \leq r$ is the cost of producing one unit of ice-cream; and (*iii*) consists of both vendors selling to their "home" market at a price of $r$, extracting the entire consumer surplus. Now suppose that prior to competing in prices, the two vendors can jointly determine the transportation cost $t$. Clearly, they will agree on some $t > (r - \phi)/2$ which allows them to sustain the cartel outcome in equilibrium.

In the interconnection model developed below there is no transportation cost in the literal sense, nor do consumers have different preferences for the firms as such. However, users call certain subsets of the population more often than others. If the majority of a consumer's preferred calling partners have chosen network $A$, and if calls from one network to the other are more expensive than calls within the same network, then this consumer derives a lower utility from subscribing to network $B$ than from joining his preferred group on $A$. In other words, a markup on outbound calls creates positive externalities among subscribers on the same network. Together with the biased calling pattern, this implies that a consumer who switches networks incurs a utility loss equivalent to the transportation cost in the ice-cream vendor model. Moreover, the interconnection fee directly affects the price differential between on-net and off-net prices, so by agreeing on a higher access markup the networks can effectively increase the perceived differentiation between themselves.

In contrast with the ice-cream vendor model, however, there are limits to the degree of differentiation that can be achieved in this way, because joining the less-preferred set of consumers is not completely without value, which implies an upper bound on the transportation cost. Moreover, the price differential has a negative impact on equilibrium profits as long as consumers make some of their calls to the other network, because a markup on off-net calls reduces total surplus and, therefore, the rents that each firm can appropriate. In terms of the ice-cream vendor model, the firms can only choose the transportation cost $t$ up to an upper bound $\bar{t}$, which defines a minimum cost of production $\underline{\phi} = r - 2\bar{t}$ above which the Nash equilibrium can be sustained through an appropriate choice of $t \leq \bar{t}$. If $\phi \geq \underline{\phi}$, then the firms maximize their equilibrium profits by choosing a transportation cost just equal to $(r - \phi)/2$.

The effects at work are quite different when access is provided at a *discount*. In this case, just as in the equilibrium found by Gans and King, calls to other networks are cheaper than those terminating on the same network, resulting in negative externalities among the customers of each firm. A network trying to expand its market share must therefore compensate the marginal consumer for the loss in utility he experiences from being on a larger network and making a smaller share of his calls at the subsidized off-net rate. A decrease in the interconnection fee implies a lower off-net rate, which increases the negative externality of network size and, thus, the compensation required to attract the marginal consumer. Consequently, the firms can increase the cost of deviating from the shared-market outcome, and hence sustain a more profitable equilibrium, by aggreeing on a reduced access charge.

While both access markups and discounts reduce the competition for retail customers, they lead to rather different effects on consumer behaviour, efficiency and profits. With a discount on interconnection and the corresponding negative externalities, consumers join the network with the smallest number of their preferred calling partners in order to maximize the number of calls priced at the more favorable off-net rate. Consequently, a large share of calls is subsidized by the firms in equilibrium, leading to inefficiently high volume demand. In constrast, when outbound calls are more expensive than on-net calls, consumers choose the same network as those whom they call the most, so the majority of calls is made at the (undistorted) on-net rate, and the main effect of the access charge is on off-equilibrium profits. For this reason, profits tend to be higher under a positive access markup when both types of equilibria exist. In particular, a markup leads to higher profits than a discount if (*i*) consumers' calling patterns are sufficiently biased towards their peer groups, (*ii*) fixed costs per customer are above a lower limit, and (*iii*) there is sufficient friction on the demand side that no firm can corner the entire market (and thereby eliminate the need to interconnect).

Regardless of whether a markup or discount is preferred by the firms, they will always agree on an access fee that is bounded away from the social optimum, so that consumers make a positive share of their calls at a price that differs from marginal cost. This implies a demand inefficiency and a loss in social welfare. Short of a regulator imposing the welfare-maximizing interconnection fee, the socially efficient outcome can be restored by a ban on termination-based price discrimination and, hence, on tariff-mediated externalities. When the price for each call is the same regardless of where it terminates, the firms strictly prefer the efficient interconnection fee as long as their profits increase in the amount of total surplus generated. This is the case, for instance, when there are subsets of captive consumers whose rents can be fully appropriated by the firms.

The model is described in the next section. Section 3.3 contains the case of an access markup. The complementary case of a discount is described in Section 3.4. The different outcomes are compared in Section 3.5. In Section 3.6 it is shown that a ban on termination-based price discrimination suffices to restore efficiency in the present model. Section 3.7 concludes. Most proofs are in the Appendix.

## 3.2 The model

The point of departure for this model is the setup in section 5 of Laffont et al. (1998b), with firms competing in non-linear tariffs and practising termination-based price discrimination. The main innovation is the introduction of biased calling patterns, i. e. a preference of consumers to call some users more often than others. Moreover, in constrast with previous research, firms are not assumed to be exogenously differentiated.

Two firms (or "networks"), $A$ and $B$, offer telephone services to end users. Both use the same technology. There is a fixed cost $f$ of connecting a customer to a network for one period. This includes the cost of billing and any other cost unrelated to variable consumption. In the case of mobile phones, the fixed cost may also include a subsidy on handsets, so that $f$ may be quite large relative to other variables. The marginal cost $c$ of each call minute includes the cost of termination $c_0$. Each network charges an access fee (or "interconnection charge") of $a$ per minute of each incoming call for the termination services which it provides at a cost of $c_0$. For the originating network, the marginal cost of an "off-net" (or "outbound") call terminating on the other network is therefore

$$\hat{c} \equiv c + a - c_0,$$

since it saves the cost of terminating the call itself but has to pay the access fee $a$. For further reference, let

$$m \equiv \frac{a - c_0}{c}$$

denote the access markup or discount relative to the true (total) marginal cost of a call, so that the perceived cost of an outbound call can be written as $(1+m)c = \hat{c}$. (Throughout, a caret denotes variables pertaining to off-net calls.)

On the demand side, there is a set of measure one of non-atomic consumers, each of whom subscribes to at most one of the two networks. All consumers make the same number of calls, also normalized to unity.[1] Each consumer's *calling pattern* determines the share of his calls going to any given subset of consumers. More precisely, let the set of all consumers be partitioned into two equal subsets (or "groups") $i = 1, 2$ of measure $1/2$ each.[2] Consumers make a share of $\gamma$ of their calls within their own group and $1-\gamma$ to the other group, with $1/2 < \gamma \leq 1$. Actual calling partners are chosen randomly in either case. See Figure 3.1 for a graphical representation.

For illustration, suppose a share of $z_1 \in [0, 1]$ of the members of group 1 is connected to network $A$ and the complementary share of $1 - z_1$ has chosen $B$.

[1]Throughout, measure-theoretic terms and the economic variables which they represent are used interchangeably.

[2]Allowing for more than two consumer groups is straightforward but does not lead to any new insights justifying the additional notation.

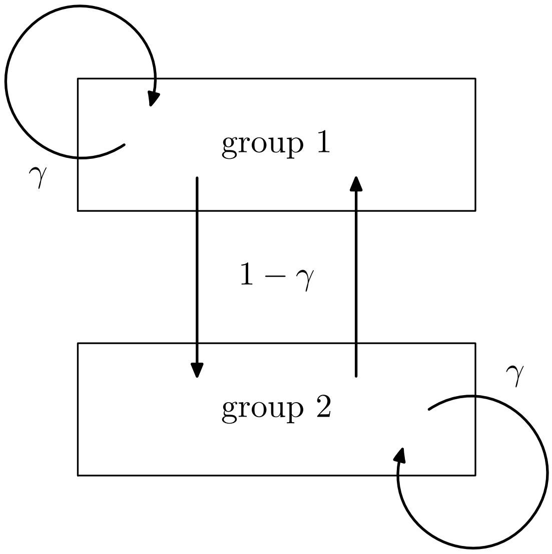


**Figure 3.1:** Consumers in each group make a share of $\gamma > 1/2$ of their calls to their own group and $1 - \gamma$ to the other.

Likewise, let $z_2$ denote the share of group 2 connected to network $A$, and let $1-z_2$ denote the share of group 2 on $B$. Then each member of group 1 makes a share of $\gamma z_1 + (1-\gamma)z_2$ of his calls to network $A$ and $\gamma(1-z_1) + (1-\gamma)(1-z_2)$ to $B$. Since $\gamma > 1/2$, consumers make a larger share of calls to their fellow group members than to the other group.[3]

In addition to belonging to different groups, consumers are heterogeneous with respect to their switching costs. A share of $\mu$ of the members of each group is free to subscribe to either network at no cost. In what follows, these consumers will be referred to as "mobile". The remaining $1-\mu$ consumers are locked-in (or "captive") to one of the networks with a switching cost of infinity (but they can leave the market if their tariff results in a negative net level of utility).[4] For

[3]Calling patterns are *biased* but still *balanced:* Consumers call some subsets of the population more often than others, but any two sets of consumers receive the same number of calls from each other. Thus, at equal usage prices, the flow of calls between the two networks is completely symmetric. See Dessein (1999a) and Dessein (1999b) for the analysis of calling patterns where this is not the case.

[4]Some form of friction is necessary for pure-strategy equilibria to exist in the presence of positive network externalities. In the literature, consumers usually vary à la Hotelling in their relative taste for each network. In such a setting, equilibrium requires a sufficiently low degree of substitutability of the two firms (i. e. a high transportation cost). In this model, in contrast, consumers regard the firms' products as a priori homogeneous, and any perception of horizontal differentiation is due entirely to prices, calling patterns and the choices made by other consumers. To obtain shared-market outcomes where both networks stay in the market, one

simplicity, all the captive members of each group are assumed to be locked-in to the same network.[5] Moreover, the captive consumers of group 1 are assumed to be locked-in on network $A$, and the captive members of group 2, on $B$. This reflects the idea that the market is "mature", i. e. the situation is essentially symmetric for both firms.

**Assumption 3** $\mu \in (1/2, 1)$.

That is, while some users are locked-in, more than half of all consumers are "mobile" and free to choose either network. This assumption eliminates some corner solutions in the presence of negative externalities; it is sufficient but not necessary for what follows.

In all other respects, consumers are identical. Their utility is linear in money, and a call of length $q$, on-net or off, increases utility by $u(q)$, where $u' > 0$ and $u'' < 0$. By assumption, consumers derive utility from calling others but not from being called.[6]

The two networks compete in two-part tariffs, and they are free to price-discriminate between on-net and outbound calls. More precisely, $T_i = (F_i, p_i, \hat{p}_i)$ is a tariff set by firm $i$ with a fixed fee $F_i$ and per-minute prices $p_i > 0$ and $\hat{p}_i > 0$ for on-net and off-net calls, respectively. Moreover, each firm can discriminate between captive and mobile consumers (but not between consumer groups).[7]

An immediate effect of termination-based price discrimination is the creation of tariff-mediated externalities among consumers choosing their network. Consumers are subject to *positive* externalities if the price of on-net calls on their network is less than the off-net rate, i. e. $p < \hat{p}$. In this case, an increase in the number of subscribers on the same network leaves each existing customer better

must rule out either firm attracting all consumers at once, eliminating the need to interconnect altogether. This restriction actually makes the model more "realistic" in the sense that real-life firms in a mature market probably cannot attract *all* their rival's customers at once by lowering their prices. (Whether in reality this is due to consumer switching costs or firms' capacity constraints is a different matter.) Another way to capture the entire market and render interconnection irrelevant would be to take over the rival firm. This is ruled out here.

[5]This assumption can be relaxed. The essential point is that the captive consumers in each group are distributed asymetrically across the two networks.

[6]This assumption, which is standard in the literature, ignores the fact that telephone calls are usually part of an underlying relationship between the caller and the party called. The main results of this paper are not affected by it.

[7]As will be shown, equilibrium tariffs for the captive and mobile consumers differ only in the fixed fee as long as termination-based price discrimination is possible. Some generality is admittedly lost by the restriction to a single two-part tariff per consumer type; the analysis of general (menues of) non-linear tariffs is beyond the scope of this paper.

off because it raises the share of his calls made at the lower on-net price. Likewise, if outbound calls are cheaper than those to the same network, consumers face *negative* externalities and experience a loss in welfare as more subscribers join their network.

The game has four stages, and the order of moves is as follows. First, in the cooperative stage, the networks jointly determine the access markup or discount, $m$. There are no side payments or binding agreements about the firms' subsequent behaviour. Second, in the competitive stage, they independently and non-cooperatively set their tariffs. Third, each consumer subscribes to exactly one network. Finally, consumers place their calls, and payoffs are realized.

The equilibrium concept is subgame-perfect Nash equilibrium. Throughout, the attention is restricted to pure strategies and symmetric "shared-market" outcomes where each firm serves half the total market (but not necessarily half of each consumer subgroup). As usual, the model is solved backwards. The last two stages are analyzed in the next two subsections. Some preliminary results for the second stage are derived in Subsection 3.2.3. Sections 3.3 and 3.4 contain, respectively, the analysis of stage 2 given an access markup and discount. The first stage is described in Section 3.5.

### 3.2.1 Stage 4: Consumption

Consider the final stage where consumers, having subscribed to one of the networks, determine their optimal consumption of telephone calls. While a consumer's calling pattern, together with the subscription choices of all other consumers, determines how many calls he makes to which network, the duration of each call is determined by its per-minute price, $p$ (or $\hat{p}$). The indirect utility that a consumer derives from a call at a per-minute price of $p$ can be expressed as

$$v(p) = u\big(q(p)\big) - pq(p),$$

where

$$q(p) = \arg\max_{q} u(q) - pq$$

is the optimal call length given $p$. By the envelope theorem, $v'(p) = -q(p)$. Moreover, $q'(p) < 0$ and $q''(p) > 0$ by assumption.

The following assumption assures that the market is viable.

**Assumption 4** $f < v(c)$.

That is, the fixed cost of connecting a consumer to a network is less than the level of utility that this consumer can achieve if all his calls are made at a per-minute price equal to the true marginal cost $c$.

### 3.2.2 Stage 3: Subscription

The subscription stage is best described as a market for network membership with unit demand and consumption externalities. The most notable feature at this stage is the discontinuity in aggregate demand induced by positive externalities. With negative externalities ($p > \hat{p}$) on both networks, subscription demand is smooth; with positive externalities on either network, the mobile consumers from the same group always choose the same network.

When consumers decide which network to subscribe to, they take the firms' tariffs and their own subsequent consumption as given. Let $z_i$, $i = 1, 2$, denote the share of consumers from group $i$ subscribing to network $A$. For an individual in group 1, the net utility of being on network $A$ is

$$\begin{aligned} w_1^A(z_1, z_2) = &\big[\gamma z_1 + (1-\gamma) z_2\big] v(p_A) \\ &+ \big[\gamma(1-z_1) + (1-\gamma)(1-z_2)\big] v(\hat{p}_A) \\ &- F_A, \end{aligned} \tag{3.1}$$

the weighted sum of the indirect utilities from on-net and off-net calls minus the fixed fee, where the weights are given by the measure of each type of calls that this consumer is planning to make. Likewise, the net utility of subscribing to network $B$ is

$$\begin{aligned} w_1^B(z_1, z_2) = &\big[\gamma(1-z_1) + (1-\gamma)(1-z_2)\big] v(p_B) \\ &+ \big[\gamma z_1 + (1-\gamma) z_2\big] v(\hat{p}_B) \\ &- F_B. \end{aligned} \tag{3.2}$$

The expressions for the members of group 2 are analogous. When on-net and off-net prices differ, the resulting externalities imply that each consumer's optimal decision depends on the choices of everyone else, giving rise to the usual multiplicity of equilibria in "network" models. The following assumption eliminates payoff-dominated equilibria that are due to coordination failure.

**Assumption 5** *If a coalition of "mobile" consumers can increase the payoff to each of its members by choosing a different network, they can coordinate on doing so. Side payments are ruled out.*

This assumption essentially prevents outcomes where the networks create positive externalities and consumers fail to choose a more favourable tariff only because they expect too few of their fellow consumers to follow suit. Thus, it makes the positive-externalities case more competitive without affecting the outcome under negative externalities, where the relevant marginal coalition always consists of an individual consumer and coordination is therefore not an issue.

The equilibrium of the subgame starting with consumers' subscription choices at stage 3, and thus the aggregate demand for network membership at stage 2, is determined by the condition that no coalition of consumers has an incentive to switch providers, given tariffs and the subscription choices made by all other users. The kind of equilibrium that can be reached depends on the type of externalities created by the firms. Consider the mobile consumers in group 1, taking the choices of all other consumers as given. If there are positive externalities ($p < \hat{p}$) on network $A$, then the incentive to choose network $A$ over $B$ increases in the number of fellow group members on $A$. If it pays for any individual mobile member of group 1 on $B$ to switch to $A$, then it is even more profitable for all of them to do so collectively. Thus, demand for network membership is discontinuous. If instead firm $A$ creates negative externalities, the marginal coalition of mobile consumers on $B$ consists of any single individual. The following lemma asserts that positive externalities on *either* network imply that demand for network membership is discontinuous on both.

**Lemma 7** *Suppose there are positive externalities (i. e. $p < \hat{p}$) on at least one of the networks. Then in any (pure-strategy Nash) equilibrium of the subgame beginning at stage 3 where consumers choose providers, all the mobile members of a given group $i$ subscribe to the same network.*

**Proof.** Without loss of generality, suppose there are positive externalities on network $A$. Let some mobile consumers from group 1 choose $A$ and some $B$. In equilibrium, they must enjoy the same level $\overline{w}$ of net utility (regardless of the type of externalities on $B$), so that no individual consumer has an incentive to deviate. But the consumers who have chosen $B$ can assure themselves a payoff $w^* > \overline{w}$ by collectively switching to $A$. Thus, firm $B$ must offer them $w^*$ in equilibrium, a contradiction. ■

The mobile consumers from the two groups $i = 1, 2$ can still choose different networks in equilibrium because they give different weights to the captive population on each network. (See Figure 3.2 for an illustration.) For the mobile consumers in group 1 to subscribe to network $A$ while those in group 2 choose $B$, the share of calls to the captive consumers from the same group, $\gamma(1-\mu)$, must be at least as large as the share of calls to the other group, $1-\gamma$. This is assured by the following assumption.[8]

**Assumption 6**

$$\gamma > \frac{1}{2-\mu}.$$

To complete the description of consumers' subscription decisions at stage 3, consider the remaining case where both firms create negative externalities, so that it is individual consumers (as against non-trivial coalitions) who have the highest incentive to switch providers. The necessary and sufficient condition for an equilibrium is the equality of the net utilities from joining networks $A$ and $B$, $w_1^A(z_1, z_2)$ and $w_1^B(z_1, z_2)$ from (3.1) and (3.2) for a mobile consumer in group 1 (and analogously for group 2), given that $z_i$ of the mobile members of group $i$ have chosen network $A$, $i = 1, 2$. Solving for market shares among each group, on obtains

$$z_1(T_A, T_B) = \frac{F_B - F_A}{\gamma\Sigma} - \frac{(1-\gamma)z_2}{\gamma} + \frac{v(\hat{p}_A) - v(p_B)}{\gamma\Sigma} \tag{3.3}$$

and

$$z_2(T_A, T_B) = \frac{F_B - F_A}{\gamma\Sigma} - \frac{(1-\gamma)z_1}{\gamma} + \frac{v(\hat{p}_A) - v(p_B)}{\gamma\Sigma} \tag{3.4}$$

where

$$\Sigma \equiv v(\hat{p}_A) - v(p_A) + v(\hat{p}_B) - v(p_B) \quad > 0,$$

and where $T_A = (F_A, p_A, \hat{p}_A)$ and $T_B = (F_B, p_B, \hat{p}_B)$ denote the tariffs offered to the non-captive consumers. Since $z_1$ is the market share of $A$ among the entire group 1, and since the locked-in members of that group always stay on network $A$, we have $z_1 \in [1-\mu, 1]$. Likewise, firm $A$ cannot attract more than the mobile members of group 2, so $z_2 \in [0, \mu]$. Since $\mu > 1/2$ by Assumption 3, the market-share equations in (3.3) and (3.4) can be solved for

$$z_1 = z_2 = z(T_A, T_B) = \frac{F_B - F_A + v(\hat{p}_A) - v(p_B)}{\Sigma}, \tag{3.5}$$

[8]There may exist (multiple) equilibria at lower values of $\gamma$ if Assumption 5 is relaxed and coordination problems are introduced.

provided that $z \in [1-\mu, \mu]$, i. e. as long as some consumers of each mobile set choose firm $A$ and some choose $B$. In this case, illustrated in Figure 3.3, the share of intra-group calls $\gamma$ is irrelevant.

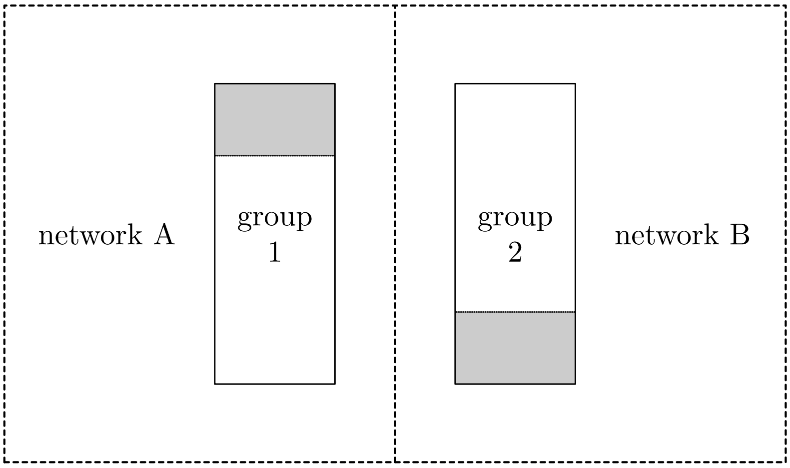


**Figure 3.2:** Consumers choose networks by groups under positive externalities. The shaded areas represent the $1-\mu$ captive consumers in each group.

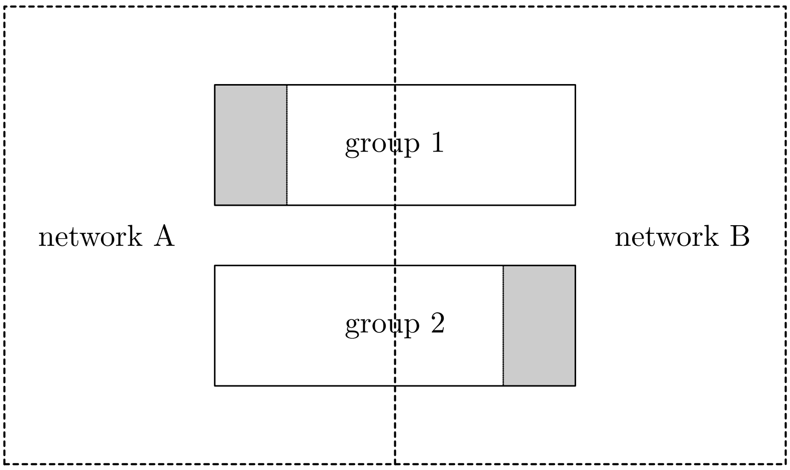


**Figure 3.3:** With negative externalities on both networks, consumers from the same group subscribe to different networks.

### 3.2.3 Stage 2: Competition in tariffs

Now consider the second stage where the networks compete in tariffs for the mobile consumers from both groups, given the access markup or discount $m$

chosen at stage 1. The following lemma gives a sufficient condition for the optimal usage prices $p$ and $\hat{p}$ to be equal to their respective marginal cost $c$ and $\hat{c}$.[9] When this condition holds, the subgame beginning at stage two is effectively reduced to a one-dimensional competition in fixed fees for market share among the mobile consumers.

**Lemma 8 (Marginal-cost pricing)** *Consider the continuation game beginning at stage 2 where the firms independently set tariffs. Suppose network $i$ offers a tariff $T_i = (F_i, p_i, \hat{p}_i)$ to a subset $X$ of all consumers. If (i) each consumer $x \in X$ choosing firm $i$ directs the same share of calls to network $A$ (and the complementary share to $B$), and (ii) tariffs are set simultaneously, then*

- *for a given set of consumers who actually join the network, the firm's profit is quasiconcave in $p$ and $\hat{p}$, and*
- *the usage prices in a pure-strategy Nash equilibrium are equal to their respective marginal cost, i. e. $p = c$ and $\hat{p} = \hat{c}$.*

The intuition is this. Given the rival's tariffs and consumers' subsequent behaviour, the firm's own tariff $T_i$ has two effects. First, it determines market shares among the mobile consumers through the net utility that each potential customer can achieve by subscribing to this firm. Second, it affects the average profit per actual customer. A necessary condition for a Nash equilibrium is that the profit per customer is maximized given the current market share, i. e. holding constant the marginal consumer's net utility. But when the average and marginal consumers make the same share of their calls off-net, then their surplus net of the fixed fee is also maximized by the same pair of usage prices $p$ and $\hat{p}$, so there is no conflict between maximizing profit per customer and satisfying the marginal consumer's incentive constraint. Setting prices equal to marginal costs, the firm maximizes the amount of surplus created, using the fixed fee to extract its share of the surplus and keep the marginal consumer indifferent.

**Corollary 1 (Marginal-cost pricing for captive consumers)** *The optimal tariff offered by each firm $i = A, B$ to its captive consumers has*

- $p = c$
- $\hat{p} = \hat{c}$

[9] Unless otherwise noted, "marginal cost" refers to the marginal cost of (one minute of) a call *as perceived by the selling firm*, i. e. $c$ for on-net calls and $\hat{c} = (1 + m)c$ for outbound calls.

- $F = \Big(\gamma z_1 + (1-\gamma) z_2\Big) v(c) + \Big(\gamma(1-z_1) + (1-\gamma)(1-z_2)\Big) v(\hat{c})$,

*where* $z_k$ *is the share of group* $k = 1, 2$ *subscribing to firm* $i$.

That is, each firm sets its usage prices such that the total surplus created in its captive market is maximized, and then uses the fixed fee to extract that surplus completely. Consequently, each firm's equilibrium profit from its captive consumers is

$$\pi_i^L = \frac{1-\mu}{2}\Big[\big(\gamma z_1 + (1-\gamma) z_2\big) v(c) + \big(\gamma(1-z_1) + (1-\gamma)(1-z_2)\big) v(\hat{c}) - f\Big], \tag{3.6}$$

regardless of the kind of tariff it offers to the mobile segment of the market.

## 3.3 Access markup and positive externalities

As shown above, consumers' subscription choices depend fundamentally on whether at least one of the firms creates positive externalities. It turns out that in the presence of an access markup, i. e. $m > 0$, a symmetric Nash equilibrium requires off-net prices above the on-net rate, i. e. $\hat{p} > p$, implying positive externalities. Conversely, when $m < 0$, then $\hat{p} < p$ in equilibrium. Thus, depending on the sign of the access markup $m$ from the first stage of the game, the firms find themselves in either of two regimes. This section describes the outcome of the stage-2 competition in tariffs given a positive access markup $m > 0$. The next section contains the analysis of the complementary case.

In a nutshell, the story in this section goes as follows. Suppose the firms agreed on a positive interconnection markup $m > 0$ at the first stage. Then the only candidate equilibrium is the symmetric outcome where both firms charge higher usage prices for off-net calls and consumers choose networks by groups. The only relevant deviation from this outcome involves one of the firms attracting all its rival's mobile consumers. This is always profitable when the access markup $m$ is small. As $m$ increases, the deviation profit decreases faster than the shared-market payoff, and under some conditions on fixed costs and calling patterns, the symmetric outcome is an equilibrium if $m$ is sufficiently large.

The following Lemma asserts that if a symmetric equilibrium exists, it must involve tariffs leading to positive externalities. Roughly speaking, profits are quasiconcave in usage prices by Lemma 8, with a maximum at $p = c$ and $\hat{p} =$

$(1+m)c$. Hence, moving closer to marginal-cost pricing leaves each firm better off, so $m > 0$ implies that the off-net price $\hat{p}$ will exceed the on-net rate $p$.

**Lemma 9** *If $m > 0$, then a symmetric equilibrium requires tariffs with $p < \hat{p}$.*

By Lemma 7, if follows that all mobile consumers in each group subscribe to the same network. It is easy to see that the mobile members of group 2 cannot prefer network $A$ if those of group 1 choose $B$. Hence, the only symmetric equilibrium candidate is the outcome where network $A$ serves all the members of group 1, and firm $B$ connects all consumers in group 2, as in Figure 3.2. By Lemma 8, usage prices must equal marginal cost in equilibrium, so the difference between on-net and off-net rates is a simple linear function of the access markup $m$.

The continuation game at stage 2 now resembles the ice-cream vendor model described in the introduction. An increase in the price of off-net calls raises the transportation cost for consumers who make more of their calls to one network than to the other. This in turn makes it less attractive for each firm to deviate from the shared-market equilibrium and attract its rival's customers. At the same time, the reservation value of subscribing to the "right" network (the equivalent of $r$ in the simple spatial model) is reduced by higher off-net prices. Thus, there is a negative effect on equilibrium profits as well as a reduced payoff from deviations. Moreover, when $m > 0$, incoming calls generate a profit for the receiving firm, and the volume of incoming and outgoing calls changes when a firm deviates from the shared-market outcome. As will be shown below, however, the overall effect of a symmetric increase in the off-net price is to reduce the lower limit for the fixed cost $f$ above which the shared-market outcome can be sustained.

Marginal-cost pricing on both networks reduces the incentive constraint for the mobile consumers in group 1 to choose $A$ to

$$F_A \leq F_B + (2\gamma - 1 - \gamma\mu)\big(v(c) - v(\hat{c})\big), \tag{3.7}$$

and likewise the constraint for the mobile members of group 2 to join network $B$ becomes

$$F_A \geq F_B - (2\gamma - 1 - \gamma\mu)\big(v(c) - v(\hat{c})\big) \tag{3.8}$$

($2\gamma - 1 - \mu\gamma > 0$ by Assumption 6). Moreover, in order for consumers' net utilities in equilibrium to be non-negative, it must be that

$$F_i \leq \gamma v(c) + (1-\gamma)v(\hat{c}) \qquad i = A, B. \tag{3.9}$$

Just as in the ice-cream vendor model, the only candidate for a symmetric pure-strategy equilibrium is the outcome where prices (in this case, fixed fees) are just equal to consumers' reservation price (the weighted indirect utility from on-net and off-net calls).

**Lemma 10** *In a symmetric (pure-strategy Nash) equilibrium, condition (3.9) must hold with equality.*

**Proof.** Consider a candidate equilibrium where the inequality in (3.9) is strict. Then there is at least one firm which can raise its fixed fee by some $\epsilon > 0$, and thus its profit per customer, without losing market share: a contradiction. ■

Taken together, Lemmas 8 and 10 imply that the profit of each firm in a shared-market Nash equilibrium with positive externalities is

$$\pi^{\oplus} = \frac{1}{2}\big(\gamma v(c) + (1-\gamma)v(\hat{c}) - f\big) + \frac{1-\gamma}{2}mcq(\hat{c}). \tag{3.10}$$

For this equilibrium to exist, neither firm must find it profitable to raise its tariff and lose its mobile clientele, or to undercut its rival and attract its customers. The first of these deviations might appear odd since it involves one of the firms voluntarily ceding the mobile market to its competitor. However, with $m > 0$ the firms make a positive profit on each incoming call which they terminate, and the amount of incoming traffic is higher when members of the same group are on different networks than in the shared-market outcome. Thus, it is not completely obvious that firm $A$ never prefers to leave the mobile market to network $B$. As the following lemma shows, however, this deviation from the candidate shared-market outcome is never profitable.

**Lemma 11** *In the shared-market outcome with alle the members of group 1 on network $A$, all group 2 consumers on network $B$, and marginal-cost usage prices $p = c$ and $\hat{p} = \hat{c}$ set by both networks, a firm ceding the mobile market to its rival reduces its payoff.*

Intuitively, since all off-net prices are equal to $\hat{c}$, the increase in incoming call volume is exactly matched by an increase in the number and volume of calls from the firm's captive consumers to all others, leaving the interconnection balance unchanged at zero. But the retreating firm loses whatever profit it makes on its mobile customers, and the surplus generated in its captive market (which it fully extracts) is reduced as a larger share of calls is made off-net.

This leaves undercutting and attracting the other firm's mobile consumers as the only remaining deviation from the candidate shared-market equilibrium. The resulting deviation profit, denoted by a superscript $D$, is[10]

$$\begin{aligned}\pi^D(T_A^D) = {} & \frac{1-\mu}{2}\Big[(\gamma + (1-\gamma)\mu)v(c) + (1-\gamma)(1-\mu)v(\hat{c}) - f\Big] \\ & + \frac{\mu}{2}\Big[(1+\mu)(p_A^D - c)q(p_A^D) + (1-\mu)(\hat{p}_A^D - \hat{c})q(\hat{p}_A^D) + 2(F_A^D - f)\Big] \\ & + \frac{1-\mu}{2}(1-\gamma+\gamma\mu)mcq(\hat{c}),\end{aligned} \tag{3.11}$$

where the first line is the profit which firm $A$ makes on its captive consumers; the second line is the profit from all mobile consumers; and the last line is the profit from terminating the incoming calls originated by the rival's captive consumers.

It turns out that this deviation is always profitable when the interconnection premium $m$ is small (but non-negative).

**Proposition 4** *There is a value $\underline{m} > 0$ such that for all $m \in [0, \underline{m}]$, it is profitable for each firm to deviate from the shared-market outcome by undercutting its rival and attracting all mobile customers, even if the deviating firm is constrained to setting usage prices equal to marginal cost, i. e. $p = c$ and $\hat{p} = \hat{c}$.*

Intuitively, when $m$ is close to zero, the deviating firm bears only a small opportunity cost in its home market (the mobile group-1 consumers for $A$) when adjusting its tariff to satisfy the incentive constraint of the rival's mobile consumers. A simple way to find a lower bound for the deviation profit is to hold prices constant at $p = c$ and $\hat{p} = \hat{c}$ and adjust only the fixed fee $F_A^D$ to compensate the marginal consumers for the loss in utility from switching networks, $(2\gamma - 1 - \gamma\mu)\big(v(c) - v(\hat{c})\big)$. But as $m$ goes to zero, so does $v(c) - v(\hat{c})$, and therefore the amount by which the deviating firm must reduce its fixed fee. As long as $f < v(c)$, as assumed, even the lower bound for the deviation profit exceeds the shared-market payoff when $m$ is close to zero.

It follows that a termination markup $m > \underline{m} > 0$ is necessary for a shared-market Nash equilibrium. What remains to be shown is whether any value of $m$ above $\underline{m}$ is *sufficient* for its existence. Since the mobile consumers in the two groups differ in their share of calls going to either network, Lemma 8 no longer applies and the optimal tariff includes usage prices deviating from marginal cost.

[10] The following equations are given for firm $A$. The expressions for firm $B$ are obtained by exchanging subscripts $A$ and $B$.

**Lemma 12 (Optimal deviation tariff)** *Suppose one of the firms departs from the shared-market outcome, attracting its rival's mobile consumers. The firm's payoff is maximized by the optimal deviation tariff* $T^D = (F^D, p^D, \hat{p}^D)$ *with*

$$
\begin{aligned}
p^D &= c - (2\gamma - 1)\frac{1-\mu}{1+\mu}\frac{q(p^D)}{q'(p^D)} \\
\hat{p}^D &= \hat{c} + (2\gamma - 1)\frac{q(\hat{p}^D)}{q'(\hat{p}^D)} \\
F^D &= (1-\gamma+\gamma\mu)v(p^D) + (1-\mu)\gamma v(\hat{p}^D)
\end{aligned}
$$

*provided that the binding incentive constraint is the one for the new customers to switch networks, i. e.* $v(p^D) - v(\hat{p}^D) \geq -\Big(v(c) - v(\hat{c})\Big)$.

Several points are worth noting about this result. First, without further restrictions on the model the optimal usage prices can be defined only implicitly. Second, $p^D > c$ and $\hat{p}^D < \hat{c}$, i. e. usage prices differ from perceived marginal cost. This is because the firm cannot price-discriminate between its marginal consumers (whom it tries to attract away from its rival) and its inframarginal ones. Satisfying the marginal consumers' incentive constraint therefore involves an opportunity cost as the reduced tariff is offered to all mobile consumers. Since the marginal consumers make more off-net calls than the average consumer, this cost is minimized by a tariff which involves a cross-subsidy from on-net to outbound calls.[11] Finally, $T^D$ has been derived under the assumption that the marginal consumer coalition consists of those mobile users who would have subscribed to the rival network in the shared-market equilibrium. It can be shown that this is correct as long as the inequality in the last line of the Lemma holds ($v(c) > v(\hat{c})$, so $p^D$ must be strictly above $\hat{p}^D$ for this condition to be violated). As the following lemma asserts, this is the case when $m$ is large enough; otherwise $\pi^D(T^D)$ constitutes an upper bound for the profit of the deviating firm.

**Lemma 13** *(i) There is a positive* $\overline{m}$ *such that* $v(p^D) - v(\hat{p}^D) \geq -\Big(v(c) - v(\hat{c})\Big)$ *if* $m \geq \overline{m}$. *(ii) When* $m < \overline{m}$, *then* $\pi^D(T^D)$ *is an upper bound for the true deviation profit.*

In a nutshell, the on-net price $p^D$ does not depend on $m$ and the off-net price $\hat{p}^D$ increases monotonically and without bound as $m$ goes to infinity. The same

[11] As $\mu$ approaches unity, $p^D$ converges to $c$ and $F^D$ to $v(c)$, and the number of calls affected by $\hat{p}^D \neq \hat{c}$ goes to zero, so in the absence of captive consumers, $T^D$ is essentially the same as the marginal-cost tariff set by the non-deviating firm.

is trivially true for $c$ and $\hat{c}$. Thus, as $m$ goes out of bounds, the LHS of the inequality is positive and the RHS negative. When $m$ is positive but below $\overline{m}$, it can be shown that the incentive constraint of the firm's existing customers must hold as well as the one for its new clients, and the maximum deviation payoff can be no higher than in the absence of this additional constraint.

The shared-market outcome is a Nash equilibrium if the payoff of each firm is non-negative and at least as large as the deviation profit. By (3.10) the derivative of the shared-market profit w. r. t. $m$ is

$$\frac{\partial \pi^{\oplus}}{\partial m} = \frac{1-\gamma}{2} m c^2 q'(\hat{c}),$$

so $\pi^{\oplus}$ is non-increasing in $m > 0$, and monotonically decreasing if $\gamma < 1$. Hence, the aim is to find the smallest positive $m$ for which deviating from the symmetric outcome is not profitable, i. e. $m^{\oplus} = \min\{m \,|\, \pi^D(T^D) \leq \pi^{\oplus}\}$.[12]

It will be convenient to analyze the effects of $m$ on the shared-market and deviation profits in terms of the fixed cost $f$. An increase in $f$ has a larger impact on the profit of the deviating firm, which connects more consumers. Thus, the condition that $\pi^D(T^D) \leq \pi^{\oplus}$ translates into a lower bound $f_{\min}$ for the fixed cost. On the other hand, the shared-market profit, too, decreases in $f$, so the non-negativity condition implies an upper bound $f^{\max}$.[13]

Comparing (3.10) with (3.11), it is immediate that the shared-market payoff is larger than the (upper bound on the) deviation profit if and only if $f$ is sufficiently large. Defining

$$\begin{aligned} f_{\min} \equiv\ & (1+\mu)(p^D - c)q(p^D) + (1-\mu)(\hat{p}^D - \hat{c})q(\hat{p}^D) + 2F^D \\ & + (1-\gamma)(1-\mu)\big(v(c) - v(\hat{c})\big) \\ & + (2\gamma - 1 - \gamma\mu) m c q(\hat{c}) \\ & - \gamma v - (1-\gamma) v(\hat{c}) \end{aligned} \tag{3.12}$$

with $T^D = (F^D, p^D, \hat{p}^D)$ from Lemma 12, we have

$$\pi^{\oplus} \geq \pi^D \quad \Longleftrightarrow \quad f \geq f_{\min}.$$

To interpret the expression in (3.12), multiply both sides by $\mu/2$, the measure of the rival's mobile clientele which the deviating firm $A$ attempts to attract. The

[12] Since $\pi^D(T^D)$ is, strictly speaking, only an upper bound on the true deviation profit (because of the possibility that $m < \overline{m}$), there might be some $m < m^{\oplus}$ that also satisfies the no-deviation condition. Thus, $\pi^{\oplus}(m^{\oplus})$ is a *lower bound* for the true shared-market profit.

[13] This corresponds to the condition that $r - 2t \leq \phi \leq r$ in the ice-cream vendor model.

first line on the RHS is the profit (gross of fixed costs) of the deviating firm from serving all mobile consumers. The second line gives the increase in $A$'s profit from increasing (and fully extracting) the surplus generated in its captive market as the locked-in users make their calls to the newly-attracted mobile consumers on-net rather than off. The third line represents the increase in its profit from incoming calls, whose volume is higher than in the shared-market outcome because the mobile and locked-in members of group 2 are now on different networks. On the other hand, the deviating firm loses the shared-market profit on the mobile group 1 consumers given by the last line, and it incurs the fixed cost $f$ on each of its newly-attracted mobile customers.

The first derivative of $f_{\min}$ w. r. t. $m$,

$$\begin{aligned}\frac{\partial f_{\min}}{\partial m} &= (1-\mu)\left[\Big((\hat{p}^D-\hat{c})q'(\hat{p}^D)+(1-2\gamma)q(\hat{p}^D)\Big)\frac{\partial \hat{p}^D}{\partial m} - cq(\hat{p}^D)\right] \\ &\quad + (1-\gamma)(1-\mu)cq(\hat{c}) \\ &\quad + (2\gamma-1-\gamma\mu)\big(cq(\hat{c})+mc^2q'(\hat{c})\big) \\ &\quad + (1-\gamma)cq(\hat{c}) \\ &= (1-\mu)\big(q(\hat{c})-q(\hat{p}^D)\big)c + (2\gamma-1-\gamma\mu)mcq'(\hat{c}),\end{aligned} \tag{3.13}$$

is negative.[14] Thus, an increase in the interconnection markup $m$ reduces the mimimum level of the fixed cost $f$ above which the shared-market outcome can be a Nash equilibrium. This is the result of a number of different effects.

First, as the interconnection markup increases, it becomes more costly for the deviating firm to satisfy the incentive constraint of the marginal consumers whom it attempts to attract away from its rival, and who (if they do switch to the deviating firm) make more of their calls off-net than the firm's average customer. A higher markup $m$ leads to a higher off-net price $\hat{p}^D$, which implies a lower fixed fee $F^D$ for all mobile consumers, marginal or not, and therefore a lower amount of surplus which the firm can extract from the mobile market. This effect, which increases the relative profitability of the shared-market outcome, is captured by the first line on the RHS of (3.13).

Second, the larger $m$, the higher is the deviating firm's profit in its captive market relative to the shared-market outcome. Prices for the locked-in consumers are always equal to marginal cost, so a higher markup on interconnection reduces the (fully extractable) surplus from off-net calls, of which there are fewer when

[14]By Lemma 12, $(\hat{p}^D-\hat{c})q'(\hat{p}^D) = (2\gamma-1)q(\hat{p}^D)$, and $\hat{p}^D < \hat{c}$, so $q(\hat{p}^D) > q(\hat{c})$.

the network is expanded. This effect, which works against the shared-market outcome, is represented by the second line of (3.13).

Third, the interconnection fee affects the profit of the deviating firm from incoming calls originated by the remaining captive consumers on the rival network. An increase in $m$ raises the profit from terminating calls but reduces their volume (because the rival network increases its off-net price $\hat{p} = \hat{c}$ accordingly). As can be seen from the third line of (3.13), the net effect is positive when $m$ is close to zero, and it converges to a non-positive value as $m$ (and, therefore, $\hat{c}$) becomes large.[15]

Finally, as the fourth line of (3.13) indicates, a larger $m$ decreases the cost of forgoing to serve the mobile group 1 consumers (for firm $A$) at the optimal shared-market tariff. In the shared-market outcome, consumers make $1-\gamma$ of their calls off-net at a usage price equal to the perceived marginal cost $\hat{c} = (1+m)c$. Thus, unless the share of intra-group calls $\gamma$ equals one, an increase in $m$ reduces the (fully extracted) surplus and therefore makes the shared-market outcome relatively less profitable.

When $\gamma$ is not too small, the first effect dominates the others as $m$ grows large, so the overall effect of an increase in $m$ is to reduce the deviation payoff relative to the shared-market profit, and thus to decrease the minimum level for the fixed cost $f$ that is necessary to sustain the symmetric equilibrium. Since by Lemmas 12 and 13, $\hat{p}^D < \hat{c}$, it can be seen from the last equation in (3.13) that $\partial f_{\min}/\partial m < 0$ if (but not only if) Assumption 6 is satisfied (cf. Figure 3.4).

Turning to the non-negativity condition for the shared-market profit, the upper bound for $f$ is given by

$$f^{\max} = v(c) - (1-\gamma)\big(v(c) - v(\hat{c}) - mcq(\hat{c})\big), \tag{3.14}$$

with a first derivative w. r. t. $m$ of

$$(1-\gamma)mc^2 q'(\hat{c}), \tag{3.15}$$

which is non-positive for all $m > 0$, and strictly negative when $\gamma < 1$: Whenever consumers make some of their calls to the other network, a markup on off-net prices reduces the amount of surplus, and therefore profits, generated in equilibrium. Consequently, shared-market profits are maximized by the smallest value of $m$ for which $f_{\min} \leq f$.

[15] $cq(\hat{c}) \to 0$ as $m \to \infty$, and $mc^2 q' < 0 \; \forall \, m > 0$.

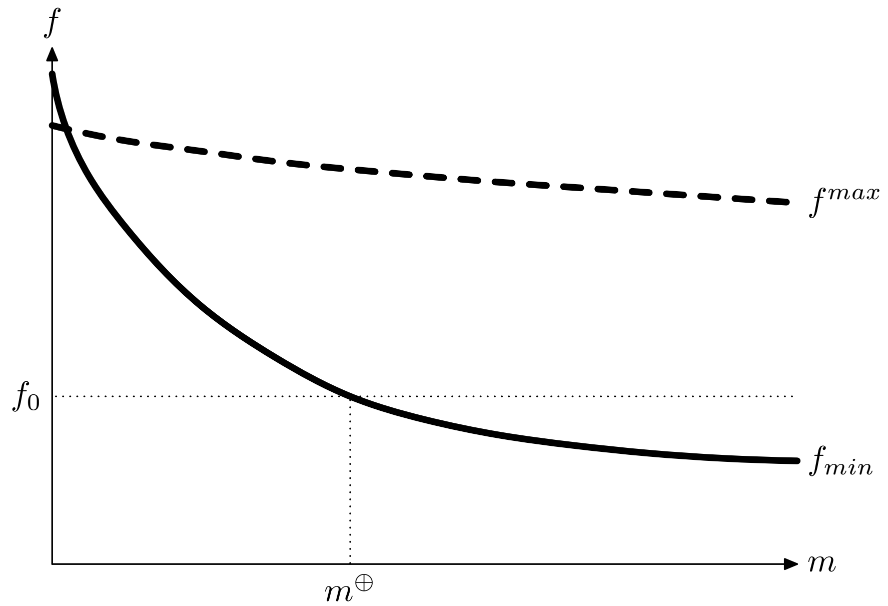


**Figure 3.4:** An example for $f_{\min}$ and $f^{\max}$ as functions of $m$. Given a fixed cost per customer of $f_0$, $m^{\oplus}$ is the smallest access markup which supports the shared-market equilibrium.

A shared-market equilibrium exists if $f_{\min} \leq f \leq f^{\max}$. The following Lemma states that when the share of intra-group calls $\gamma$ is large, then $f_{\min}$ and $f^{\max}$ define a non-empty interval for the fixed cost $f$, as in Figure 3.4.

**Lemma 14** *If $m \geq \overline{m}$ from Lemma 13, then there is a $\underline{\gamma} < 1$ such that $f_{min} \leq f^{max}\ \forall \gamma > \underline{\gamma}$.*

Moreover, the lower the share of mobile consumers $\mu$, the wider is the range of parameters over which the shared-market equilibrium can be sustained. Unless $m$ is too close to zero, the threshold $f_{\min}$ for the fixed cost $f$ increases in $\mu$. Lower values of $\mu$ imply a smaller set of consumers which a deviating firm can attract, and a larger share of off-net calls made by each of them, making their incentive constraint more costly to satisfy.[16] Since this reduces the deviation profit, a lower fixed cost $f$ suffices to support the shared-market outcome when $\mu$ is exogenously decreased.

Thus, when the calling pattern of each group is sufficiently biased towards the members of the same group, and when there is some degree of friction in the market, there is a range of values for the (exogenous) fixed cost $f$ such that $m$

[16]The last part of this statement relies on the assumption that $\mu > 1/2$.

can be chosen to sustain the shared-market outcome in equilibrium with positive profits.

**Proposition 5** *Suppose the share of intra-group calls* $\gamma$ *is sufficiently high and the fixed cost exceeds the lower bound of* $f_{min}$,

$$f > \lim_{m \to \infty} f_{min}(m).$$

*If the interconnection markup* $m$ *has been chosen at stage 1 such that*

$$f \geq f_{min}(m),$$

*then the unique Nash equilibrium in pure strategies is the shared-market outcome with marginal-cost usage prices* $p = c$ *and* $\hat{p} = \hat{c}$*, consumers choosing networks by group as in Figure 3.2, and per-firm profits as in (3.10).*

Whether or not the lower bound of $f_{\min}$ lies below the fixed cost $f$ depends on the exact functional specification. It is not difficult, however, to construct examples where $f_{\min}$ is negative above some finite value $m$, in which case the shared-market outcome can always be sustained if $\gamma$ is sufficiently high.

## 3.4 Access discount and negative externalities

This section covers the complementary case where the firms have agreed on an interconnection discount $m < 0$ at stage 1.[17] As in the previous section, there is no shared-market equilibrium where externalities and the access markup $m$ are of opposite signs; if they were, then one of the firms could increase its profit by approaching marginal-cost pricing while holding market shares constant.

**Lemma 15** *When* $m < 0$*, a symmetric Nash equilibrium requires* $\hat{p} < p$.

With negative externalities on both networks (lower prices for off-net calls than on-net), the market shares of firm $A$ among the two groups are given by $z(T_A, T_B)$ from (3.5) as long as each network is chosen by some members of each subset, i. e. while $z \in [1-\mu, \mu]$. When this is the case, all a firm's mobile customers make the same share $z$ of their calls to network $A$ and $1-z$ to $B$, so

[17] This case is included mainly for the sake of completeness, and because it plays a role in the literature. In reality, access discounts and off-net prices below the on-net rate appear to be less relevant, although there are examples of "bill and keep" arrangements with an access charge of zero; cf. Gans and King (1999).

by Lemma 8 usage prices in a Nash equilibrium must be $p = c$ and $\hat{p} = \hat{c}$. This means that the firms compete in fixed fees $F_A$ and $F_B$ only, and the expression for the market share of network $A$ is reduced to

$$z(F_A, F_B) = \frac{1}{2} + \frac{F_B - F_A}{2\Delta},$$

where

$$\Delta \equiv v(\hat{c}) - v(c) > 0$$

is the difference in indirect utilities from outbound and on-net calls priced at their respective marginal cost.

Now consider the profit of firm $A$ given marginal-cost pricing by both firms,

$$\begin{aligned}\pi_A(F_A, F_B) =& \frac{1-\mu}{2}\left[zv(c) + (1-z)v(\hat{c}) - f\right] \\ &+ \left(z - \frac{1-\mu}{2}\right)(F_A - f) \\ &+ z(1-z)mcq(\hat{c}),\end{aligned} \tag{3.16}$$

where the first line on the RHS is the profit $\pi_A^L$ that firm $A$ makes on its captive consumers, as in (3.6) with $z_1 = z_2 = z$; the second line is the profit from the mobile subscribers whom $A$ attracts; and the last line gives the (negative) profit from incoming calls, whose volume is a function of firm $B$'s off-net prices for its mobile and captive customers, both of which are equal to $\hat{c}$. (The arguments of $z$ have been dropped for simplicity.) Since the net utility of the mobile consumers must be non-negative, the fixed fee is constrained by

$$F_A \leq \frac{v(c) + v(\hat{c})}{2}. \tag{3.17}$$

As long as this constraint is slack, the first-order condition w.r.t. $F_A$ is given by

$$\begin{aligned}\frac{\partial \pi_A}{\partial F_A} =& \frac{1-\mu}{4} \\ &+ \frac{\mu}{2} + \frac{F_B + f - 2F_A}{2\Delta} \\ &+ \frac{F_B - F_A}{2\Delta^2} mcq(\hat{c}),\end{aligned} \tag{3.18}$$

which yields $A$'s best-reply function, implicitly defined by

$$\big(2\Delta + mcq(\hat{c})\big)F_A = \big(\Delta + mcq(\hat{c})\big)F_B + \Delta f + \frac{1+\mu}{2}\Delta^2. \tag{3.19}$$

The best-reply function for $B$ is completely symmetric, i. e.

$$\big(2\Delta + mcq(\hat{c})\big)F_B = \big(\Delta + mcq(\hat{c})\big)F_A + \Delta f + \frac{1+\mu}{2}\Delta^2. \tag{3.20}$$

The objective function of $A$ in (3.16) is concave in $F_A$ if $2\Delta + mc\hat{q}$ is positive:

$$\frac{\partial^2 \pi_A}{\partial F_A^2} = -\frac{1}{2\Delta^2}\big(2\Delta + mcq(\hat{c})\big).$$

The following Lemma assures that this is the case when $m$ is negative but not too large in absolute terms.

**Lemma 16** *Consider the function $g(m) = (1+y)\Delta + mcq(\hat{c})$, where $\Delta \equiv v(\hat{c}) - v(c)$. If $y = 0$, then $g(m) < 0\ \forall m < 0$. If $y > 0$, then there is a value of $\tilde{m} < 0$ such that $g(m) > 0\ \forall m \in (\tilde{m}, 0)$.*

In particular, this means that $\Delta + mcq(\hat{c})$ is negative whenever $m$ is negative, and $2\Delta + mcq(\hat{c})$ is positive when $m$ is "not too negative". ($2\Delta + mcq(\hat{c})$ is concave in $m$ for all $m < 0$.) Hence, the profit function is concave and (3.19) and (3.20) are indeed best-reply functions, and their slope is negative. This implies that fixed fees are strategic *substitutes*. As can be seen from (3.18), the cross-partial derivative of firm $A$'s profit w.r.t. $F_B$,

$$\frac{\partial^2 \pi_A}{\partial F_A \partial F_B} = \frac{1}{2\Delta^2}\big(\Delta + mcq(\hat{c})\big),$$

is indeed negative. More precisely, an increase in $F_B$ has two opposite effects. First, as usual, a higher fixed fee $F_B$ makes raising $F_A$ more profitable for firm $A$; this is reflected by the positive term $\Delta$ in the parenthesis. Second, there is a negative effect of an increased $F_B$ represented by the term $mcq(\hat{c})$. When $m < 0$, the profit from terminating incoming calls (given by the last line in (3.16)) is a negative, strictly convex function of $F_A$ with a minimum at $F_A = F_B$ (i. e. at $z = 1/2$). Therefore, the effect on $A$'s profit of an increase in $F_A$ increases monotonically as $F_A - F_B$ increases, so raising $F_B$ makes an increase in $F_A$ less profitable. By Lemma 16 this second, negative effect dominates the first whenever $m$ is negative.

The best-reply functions for $A$ and $B$ define a unique and symmetric pure-strategy candidate equilibrium in $F_A$ and $F_B$, which is stable if $3\Delta + 2mcq(\hat{c}) > 0$ (so that the slopes of the best-reply functions are larger than $-1$). Solving (3.19) and (3.20) for $F_A = F_B = F^{\ominus}$, one obtains

$$F^{\ominus} = f + \frac{1+\mu}{2}\Delta. \tag{3.21}$$

Together with the marginal-cost pricing result for the usage prices and the constraint on the fixed fee in (3.17), this leads to the following characterization of the shared-market outcome.

**Proposition 6** *In the shared-market outcome with negative externalities, usage prices are equal to marginal cost,*

$$p = c \quad \textit{and} \quad \hat{p} = \hat{c},$$

*fixed fees exceed fixed costs,*

$$F_A = F_B = F^{\ominus} = \min \left[ f + \frac{1+\mu}{2}\Delta, \frac{v(c)+v(\hat{c})}{2} \right],$$

*and each firm serves half of each consumer group, $z_1 = z_2 = 1/2$, as in Figure 3.3.*

It follows that equilibrium profits per firm are

$$\pi^{\ominus}(m) = \frac{1-\mu}{2}\left(\frac{v(c)+v(\hat{c})}{2} - f\right) + \frac{\mu}{2}\left(F^{\ominus} - f\right) + \frac{mcq(\hat{c})}{4} \tag{3.22}$$

where the first term gives the equilibrium profit on firm $i$'s captive consumers, the second term is the profit from non-captive consumers, and the last term gives the loss from granting access below cost.[18]

**Proposition 7** *In the symmetric, shared-market outcome with negative externalities, a decrease in the access markup $m$ increases profits if $m$ is negative but close to zero.*

**Proof.** Analogous to the proof of Lemma 16. For $m \approx 0$, $F^{\ominus} = f + \frac{1+\mu}{2}\Delta$, and the derivative of the profit per firm w.r.t. $m$ is

$$\frac{\partial \pi^{\ominus}}{\partial m} = \frac{c}{4}\Big(mcq'(\hat{c}) - \mu^2 q(\hat{c})\Big). \tag{3.23}$$

As $m \to 0$, the first term in the parenthesis vanishes, and the expression is negative. Thus, $\pi^{\ominus}$ decreases as $m$ approaches zero from below. ■

[18]To verify that this candidate outcome is indeed a Nash equilibrium, it remains to be shown that it is robust against unilateral deviations outside the range of $z \in [1-\mu, \mu]$, as well as against either firm deviating to a tariff with $p < \hat{p}$. Non-existence of equilibrium under some parameter constellations would still imply that (3.22) is an *upper bound* on shared-market profits under an interconnection discount.

It follows that profits in the shared-market outcome with negative externalities are maximized by some $m^{\ominus}$ strictly below zero.[19] Intuitively, the lower is $m$, the larger is the difference $p - \hat{p}$ between on-net and off-net rates, so the stronger are the negative externalities that consumers on the same network exert on each other. When a firm tries to expand its market share, the marginal consumer has to be compensated for the loss in utility resulting from being on a larger network; the larger is $p - \hat{p}$, the higher is the required compensation (which is given to all inframarginal consumers as well). By making it more expensive to expand, a lower $m$ helps to sustain a more profitable shared-market outcome.

However, there is a lower bound for the optimal $m^{\ominus}$. First of all, a discount on incoming calls on both networks leads to off-net prices below their true marginal cost, i. e. $\hat{p} = (1 + m)c < c$ for all $m < 0$. Even if access payments between the two networks cancel out in equilibrium, each firm still has to subsidize the incoming calls originated by its rival's customers. This subsidy, given by the last term in (3.22), has an increasingly negative effect on profits as $m$ decreases, because both the subsidy per call minute $mc$ and the length of incoming calls $q((1+m)c)$ increases. At what point this overproduction effect dominates the competition effect depends on functional forms and parameters; but while the negative externalities induced by $m < 0$ soften the competition for market share, it is clear that they also distort *equilibrium* consumption, and hence reduce profits.

Second, $m^{\ominus}$ is bounded below by the condition that $3\Delta + 2mcq(\hat{c}) > 0$, which is necessary for the slopes of the reaction functions to exceed $-1$. (The same condition implies that $2\Delta + mcq(\hat{c}) > 0$, guaranteeing the concavity of the objective function.)

Third, as the following lemma asserts, the mobile consumers' participation

[19] This is essentially the same result as in the analysis in Gans and King (1999) of the model in Laffont et al. (1998b) with two-part tariffs and termination-based price discrimination. In that model, consumers' utility is specified explicitly as

$$u(q) = \frac{q^{1-(1/\eta)}}{1 - \frac{1}{\eta}},$$

with $\eta > 1$. It follows that variable demand is given by $q(p) = p^{-\eta}$, so that $q(\hat{c}) = \hat{c}^{-\eta}$ and $q'(\hat{c}) = -\eta\hat{c}^{-(\eta+1)}$. Setting the first derivative in (3.23) equal to zero and solving for $m^{\ominus}$, one can see that profits in the shared-market outcome with negative externalities are maximized by

$$m^{\ominus} = -\frac{\mu^2}{\eta + \mu^2} \quad \in (-1, 0).$$

constraint in (3.17) is binding whenever $m$ is below some negative cutoff value. When this is the case, profits *increase* in $m$, implying another lower bound for the optimal $m^{\ominus}$.

**Lemma 17** *Given the fixed fee $F^{\ominus}$ from (3.21), the individual-rationality constraint in (3.17) is binding if and only if $m < m^{PC}$, where $m^{PC} \in (-\infty, 0)$ solves*

$$f + \frac{1+\mu}{2}\Delta = \frac{v(c) + v(\hat{c})}{2}.$$

*Moreover, if $m < m^{PC}$, then $\partial \pi^{\ominus} / \partial m > 0$.*

## 3.5 Stage 1: Cooperative choice of the access charge

When the networks jointly determine the interconnection markup at the first stage of the game, they either choose a discount ($m^{\ominus} < 0$) to support a shared-market outcome with negative externalities, as in (3.22); or they can agree on a positive markup $m^{\oplus} > 0$ in order to sustain the equilibrium with positive externalities, with profits per firm as in (3.10). In either case, Propositions 4 and 7 imply that the markup $m$ chosen by the networks is bounded away from the socially optimal value of zero.

An important difference between the two types of externalities is their relative effect on payoffs in and off the equilibrium. With negative externalities induced by an access discount, consumers choose networks so as to maximize their number of off-net calls. This also maximizes the equilibrium share of calls with a usage price below the true marginal cost $c$, and thus with inefficiently high volume demand. In contrast, in the positive-externalities outcome described in Section 3.3, the access charge mainly affects payoffs *off* the equilibrium. It turns out that the positive-externalities outcome is more attractive to the firms if the share of intra-group calls $\gamma$ and the fixed cost $f$ are both sufficiently high.

When the firms agree on an access markup, the shared-market outcome with consumers joining networks by group can be sustained if the fixed cost exceeds the lower bound of $f_{\min}$. Since equilibrium profits decrease in $m$, the profit-maximizing $m^{\oplus}$ is chosen such that $f_{\min}$ (a monotonically-decreasing function of $m$) just equals the actual fixed cost $f$. By (3.10), per-firm profit $\pi^{\oplus}$ is continuous

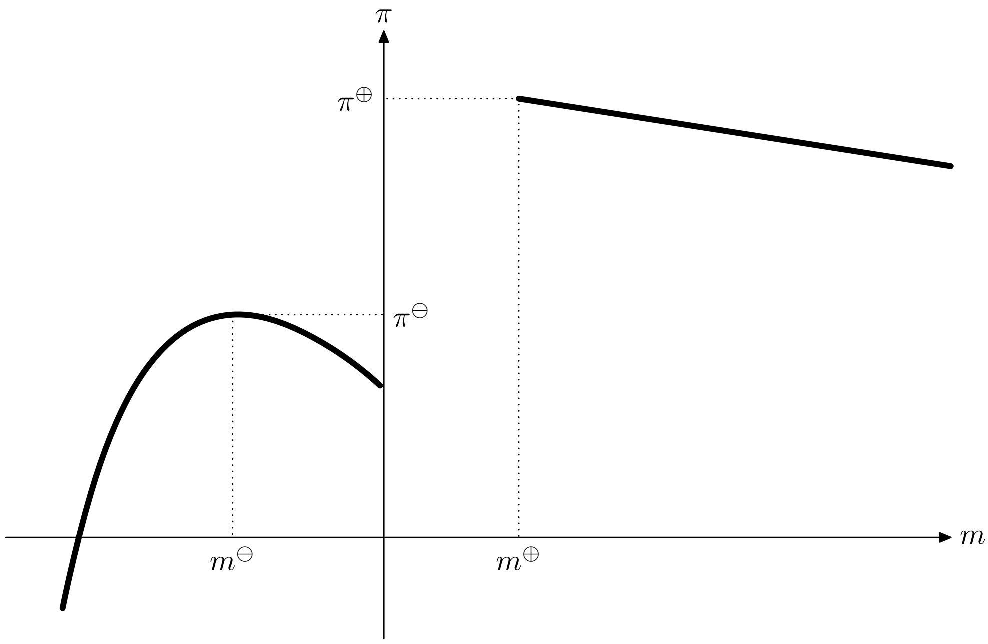


**Figure 3.5:** Equilibrium shared-market profits as a function of the interconnection markup $m$ when $\gamma$ and $f$ are sufficiently high for $\pi^{\oplus}$ to exceed $\pi^{\ominus}$. No equilibrium exists for $m \in [0, m^{\oplus})$.

and monotonically increasing in the share of intra-group calls $\gamma$. As $\gamma$ approaches unity, $\pi^{\oplus}$ converges to the cartel level,

$$\gamma \to 1 \quad \Rightarrow \quad \pi^{\oplus} \to \frac{v(c) - f}{2}.$$

Not only do profits increase in $\gamma$, but the existence of the positive-externalities equilibrium also requires that $\gamma > 1/(2-\mu)$, as assumed.[20] In contrast, as $\gamma$ goes to 1/2, so that the calling pattern becomes "unbiased" in the sense of Laffont et al. (1998a), there is no shared-market equilibrium with positive externalities.

On the other hand, if the firms negotiate an access discount and subsequently offer tariffs which lead to negative externalities, then by (3.22) equilibrium shared-market profits are independent of $\gamma$. In particular, this is true for the upper bound on $\pi^{\ominus}$,

$$\frac{1}{2}\left(\frac{v(c) + v(\hat{c}) + mcq(\hat{c})}{2} - f\right).$$

[20]See Assumption 6. The lower bound for $\gamma$ of $1/(2-\mu)$ is a consequence of Assumption 5, which rules out coordination failure among consumers. To the extent that consumers cannot coordinate their choice of network, there may also exist (multiple) positive-externalities equilibria for lower values of $\gamma$.

Moreover, Lemma 16 implies that this upper bound is strictly less than $\big(v(c) - f\big)/2$ for all $m < 0$, including the profit-maximizing $m^{\ominus}$ from the previous section.[21] Proposition 8 follows directly.

**Proposition 8** *Suppose the fixed cost $f$ is sufficiently high to support the positive-externalities equilibrium. Then there is a minimum value $\gamma^* < 1$ for the share of intra-group calls $\gamma$ such that equilibrium profits are maximized by a positive interconnection markup $m^{\oplus} > 0$ whenever $\gamma \geq \gamma^*$.*

Whether $f$ is indeed sufficiently high to support the positive-externalities equilibrium depends in part on the share of mobile consumers $\mu$. The larger is $\mu$, the larger must be the share of intra-group calls $\gamma$ to support the outcome depicted in Figure 3.2, which requires that $\gamma > 1/(2-\mu)$. Moreover, it is relatively more profitable to deviate and attract the rival's subscribers when the share of mobile consumers is large. Indeed, as $\mu \to 1$, the positive-markup outcome cannot be sustained by any $f < f^{\max}$. Thus, a certain amount of friction (measured by $1-\mu$) is necessary to support the shared-market equilibrium with a positive markup on interconnection and on outbound calls. In contrast, the negative-externalities equilibrium exists even when all consumers are mobile.

## 3.6 Banning termination-based price discrimination

Since the profit-maximizing access markup or discount is bounded away from zero, social welfare is compromised in either case because some calls are priced above or below their true cost, implying a demand inefficiency. (The exception is the limit case under positive externalities where $\gamma = 1$, so that there are no calls between the networks; but this exploits the assumption that all the locked-in members of each group are on the same network.) This prompts the question of what a regulator can and should do to improve the outcome. In a world of complete information the answer would be simple: Impose $m = 0$ and let the networks compete from stage 2. In reality, regulation introduces a set of problems of its own, most notably asymmetric information between the firms and the regulator and the risk of regulatory capture. It turns out that in the present setup, there is no need for the regulator to determine the correct access charge. Instead, a

[21] Since $F^{\ominus} \leq \big(v(c) + v(\hat{c})\big)/2$ by Proposition 6, the upper bound for profits follows from (3.22).

ban on termination-based price discrimination suffices for the networks to strictly prefer an access markup of zero. This result goes beyond the profit neutrality of the interconnection fee generally found in the literature on non-linear tariffs.

When on-net and off-net prices must be the same, then by definition the level of net utility from joining either network is independent of the decisions taken by all other consumers. Consider the set of all mobile consumers (from both groups) and let $p_i$ denote the usage price set by network $i = A, B$ for all calls, and $F_i$ the fixed fee. If $v(p_A) - F_A > v(p_B) - F_B$, then all mobile consumers join network $A$. If the inequality is reversed, they all choose $B$. For the remaining case where $v(p_A) - F_A = v(p_B) - F_B$ and users are indifferent between the two networks, it is assumed that each consumer independently chooses $A$ or $B$ with equal probability (cf. Figure 3.6).[22]

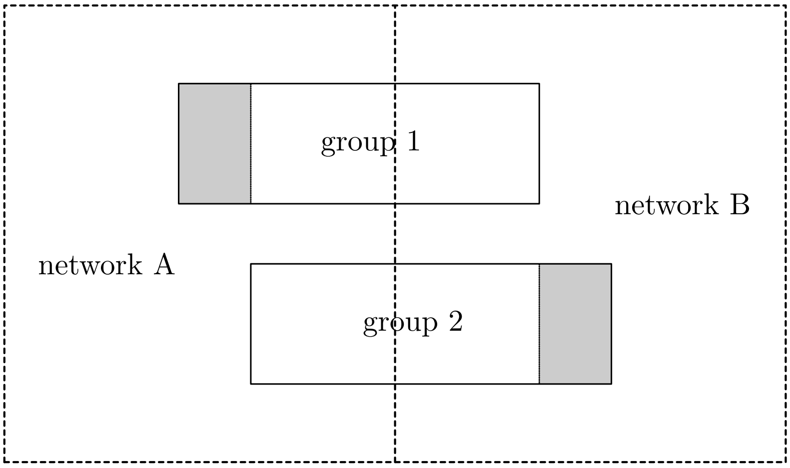


**Figure 3.6:** In the absence of externalities, mobile consumers independently choose one of the networks with probability 1/2 if both offer the same level of net utility $v(p_i) - F_i$.

Thus, there are only two equilibrium candidates. In the symmetric outcome, each network serves its captive consumers plus one-half of the mobile consumers

[22] This assumption, which is admittedly arbitrary, is sufficient but not necessary for what follows. Cf. footnotes 23 and 24.

from each group and makes a "shared-market" profit of

$$\begin{aligned}\pi_i^S =&\frac{1-\mu}{2}\Big((p_i^L-c)q(p_i^L)+v(p_i^L)-f\Big)+\frac{\mu}{2}\Big((p_i^M-c)q(p_i^M)+F_i^M-f\Big)\\&+\frac{1-\mu}{2}\Big(1-\frac{\mu}{2}-\gamma(1-\mu)\Big)\Big[q(p_j^L)-q(p_i^L)\Big]\\&+\frac{\mu}{4}\Big[q(p_j^M)-q(p_i^M)\Big]\end{aligned}\tag{3.24}$$

where the subscript $j$ denotes the rival's tariff, and superscripts $L$ and $M$ denote locked-in and mobile consumers, respectively. The first line of (3.24) gives the direct profits from captive and mobile consumers; the remaining terms represent the access payments to and from the rival network. Thus, (*i*) the optimal usage prices in the shared-market outcome must be the same on both networks, (*ii*) equilibrium prices are equal to their average marginal cost, i. e. $p_i^{LS}=c+\big(1-(1-\mu)\gamma-\mu/2\big)mc$ and $p_i^{MS}=c+mc/2$, and (*iii*) profits are quasi-concave in usage prices. Moreover, access payments cancel out in equilibrium, so the terms on the second and third line of (3.24) are zero.

To corner the (mobile) market, network $i$ only needs to provide a level of net utility $v(p_i^M)-F_i^M$ exceeding by some $\epsilon>0$ that which is offered by the rival network $j$. The resulting profit for firm $i$ is

$$\begin{aligned}\pi_i^K =&\frac{1-\mu}{2}\Big((p_i^L-c)q(p_i^L)+v(p_i^L)-f\Big)+\mu\Big((p_i^M-c)q(p_i^M)+F_i^M-f\Big)\\&-\frac{1-\mu}{2}mc\Big[(1-\gamma)(1-\mu)q(p_i^L)+\mu q(p_i^M)-(1-\gamma+\gamma\mu)q(p_j^L)\Big].\end{aligned}\tag{3.25}$$

The optimal usage prices maximizing $\pi_i^K$ are again equal to the average marginal cost of each call. Thus, the firm sets a price of $p_i^{LK}=c+(1-\gamma)(1-\mu)mc$ for its captive consumers, and a price of $p_i^{MK}=c+\frac{1-\mu}{2}mc$ for its mobile customers.

**Lemma 18** *Suppose $m\neq 0$. If a symmetric (shared-market) Nash equilibrium exists, then per-firm profits must be strictly less than*

$$\frac{1-\mu}{2}\big(v(c)-f\big).$$

**Proof.** In a symmetric shared-market equilibrium, access payments between the networks cancel out, so the terms on the second and third line of (3.24) are zero. The profit on each firm's locked-in consumers given by the first term on the first line is strictly less than

$$\frac{1-\mu}{2}(v(c)-f)$$

because

$$(p-c)q(p)+v(p)$$

is uniquely maximized by $p=c$, and $p_i^{LS}\neq c\ \forall m\neq 0$.

To prove the lemma, it suffices to show that the profit made on the firm's mobile clientele,

$$\frac{\mu}{2}\Big((p^{MS}-c)q(p^{MS})+F^{MS}-f\Big), \tag{3.26}$$

must be non-positive in equilibrium. Otherwise, the following deviation from the shared-market outcome would be profitable. Let one of the firms reduce its fixed fee $F^{MS}$ by some $\epsilon>0$, cornering the market while keeping all other prices constant. By (3.24) and (3.25), this results in a net gain of

$$\frac{\mu}{2}\Big((p^{MS}-c)q(p^{MS})+F^{MS}-f-\epsilon\Big)+\frac{1-\mu}{2}mc\mu\Big[q(p^{LS})-q(p^{MS})\Big].$$

Since $p^{LS}<p^{MS}$ for $m>0$, and the reverse for $m<0$, there always exists some $\epsilon>0$ such that this deviation is profitable if (3.26) is positive.[23] ■

The only remaining candidate equilibrium is asymmetric, with one firm cornering the mobile market and the other serving only its locked-in consumers. The profit of the firm which retreats to its captive market, denoted by a superscript $R$,

$$\begin{aligned}\pi_i^R=&\frac{1-\mu}{2}\Big((p_i^L-c)q(p_i^L)+v(p_i^L)-f\Big)\\&+\frac{1-\mu}{2}mc\Big[(1-\gamma)(1-\mu)q(p_j^L)+\mu q(p_j^M)-(1-\gamma+\gamma\mu)q(p_i^L)\Big],\end{aligned} \tag{3.27}$$

is maximized by a usage price of $p_i^{LR}=c+(1-\gamma+\gamma\mu)mc$, which is again equal to average marginal cost.

**Lemma 19** *If $m\neq 0$, then there is no asymmetric Nash equilibrium in (weakly) undominated pure strategies.*

[23] It can be shown that $|p^{LS}-c|\leq|p^{MS}-c|$ (which is sufficient but not necessary for the proof) whenever at least half of group 1 (2) joins network $A$ ($B$). Thus, the assumption that indifferent consumers choose networks with equal probability is sufficient but not necessary.

**Proof.** See Appendix.[24]

An access markup of zero maximizes both profits and welfare. Optimal usage prices for all firms converge to $c$ as $m$ goes to zero, and the profit of each firm in its captive market becomes $(1-\mu)\big(v(c)-f\big)/2$. Moreover, equilibrium in the mobile market requires that the profit per mobile consumer, $F^M - f$, be zero. There are multiple equilibria, but they only differ in which firm serves which part of the mobile market at zero profit.

**Proposition 9** *The unique profit-maximizing interconnection charge is equal to cost, i. e. $m = 0$. In equilibrium, usage prices are equal to (technical) marginal cost, $p = c$, and each firm achieves a profit of*

$$\frac{1-\mu}{2}\big(v(c)-f\big).$$

This is the first-best outcome. Both firms have a strict incentive to set $m = 0$ to the best of their knowledge, and welfare is maximized since prices are equal to the true marginal cost of each call. The captive consumers receive no rent at all, while the mobile users extract the entire surplus generated in their market segment.

This result differs from most of the literature in that the firms competing in nonlinear tariffs strictly prefer to choose the welfare-maximizing interconnection fee. In contrast, Laffont et al. (1998a) and Hahn (2000) find that the access charge has no effect on equilibrium profits in a model with non-linear tariffs. Dessein (2000) also obtains the profit-neutrality result as long as all consumers participate in the market, so that aggregate demand for network membership is constant. This is due to the fact that in the Hotelling model with exogenous differentiation employed in those papers, equilibrium profits depend only on the transportation cost, leaving the firms indifferent about marginal cost and, hence, interconnection fees.

In the present model, a ban on termination-based price discrimination prevents the networks from differentiating themselves and, therefore, from achieving a positive profit in the mobile market. At the same time, a deviation from the socially-optimal access charge reduces the amount of surplus that each firm can

[24] The proof relies on Assumption 3 ($\mu > 1/2$) and on the assumption than when indifferent, each consumer joins either network with equal probability. Both assumptions together are sufficient to prove the lemma, but neither one is necessary.

extract from its captive customers.[25] However, it should be kept in mind that the type of calling patterns modelled here is still quite special. Some scope for collusion might well be restored if calling patterns are unbalanced and consumers differ in their willingness to switch networks, as in Dessein (1999a) and Dessein (1999b).

## 3.7 Conclusion

This paper has been focussed on a mature telecommunications market where two networks compete for retail customers but cooperate when setting a reciprocal interconnection fee. When the firms can price-discriminate based on where a call is terminated, then a negotiated access charge can be used to sustain collusion in the final market. An access markup or discount implies that in a shared-market equilibrium consumers are subject to positive or negative externalities. In either case, competition for market share is softened. In particular, when (*i*) consumers make a sufficiently high share of calls within their own group, (*ii*) some consumers are locked in, and (*iii*) the fixed cost per customer is sufficiently high, then a positive access markup can be used to sustain a shared-market equilibrium with positive externalities and profits close to the cartel level.

In contrast with previous research, this prediction appears to be roughly in line with casually-observed data from the mobile-telephony markets in Germany, the UK and other countries, where access charges appear to exceed quite significantly the marginal cost of interconnection, and where off-net calls are generally more expensive than on-net calls.

From a theoretical point of view, the main innovation of the model is to endogenize the horizontal differentiation between the networks, based on a plausible assumption about consumer preferences. This is in contrast with much of the existing literature based on address models à la Hotelling, where equilibrium profits under non-linear pricing are determined entirely by the exogenous transportation cost. Moreover, the results presented here are derived under a functional specification which, while being far from general, is not very restrictive by the standards of existing models.

As far as policy is concerned, the results in this paper suggest that termination-based price discrimination should be viewed with suspicion when

[25] Dessein (2000) obtains a similar result in a discrete-choice model with elastic total demand for network membership.

the networks compete in non-linear tariffs. (With linear tariffs, price discrimination can actually increase welfare; see Laffont et al. (1998b)). Most importantly, banning this form of price discrimination can be an efficient substitute for a more difficult direct regulation of the access charge, at least within the confines of this model.

Finally, the results suggest that number portability might have an effect on competition beyond the obvious reduction of consumer switching costs. To the extent that termination-based pricing schemes rely on number prefixes to identify network membership, number portability might render this form of price discrimination impracticable, eliminating the incentive to deviate from the socially efficient interconnection fee.

## 3.8 Appendix

**Proof of Lemma 8**

In any candidate Nash equilibrium where firm $i$ offers a tariff $T_i$ to the consumers in $X$, it attracts a (possibly empty) subset $Z \subseteq X$. Let $z \in Z$ denote a consumer in $Z$, and let $a(z)$ denote the share of his calls made "on-net" on network $i$. The incentive constraint of the marginal consumer $z^m$ (i. e. the individual or coalition who has the highest incentive to deviate to network $j$) is given by

$$a(z^m)v(p_i) + (1 - a(z^m))v(\hat{p}_i) - F_i \geq w_j,$$

where $w_j$ is the net utility which the marginal consumer (or coalition) can assure himself by switching to network $j$. Note that $w_j$ is exogenous for firm $i$. An optimal tariff $T_i^*$ must maximize firm $i$'s profit subject to the above incentive constraint. Precisely,

$$\begin{aligned} T_i^* = \arg\max \int_{z \in Z} & a(z)(p_i - c)q(p_i) + (1 - a(z))(\hat{p}_i - \hat{c})q(\hat{p}_i) + F_i - f dz \\ & + I(Z, T_j) - \lambda\Big[w_j + F_i - a(z^m)v(p_i) - (1 - a(z^m))v(\hat{p}_i)\Big] \\ & + \pi_i^R(Z, T_i^R), \end{aligned}$$

where $\lambda$ is the usual Kuhn-Tucker shadow price of the constraint, $I(Z, T_j)$ is the profit from incoming calls, and $\pi_i^R(Z, T_i^R)$ is the profit on firm $i$'s remaining customers who are offered a tariff $T_i^R$. Note that $I(\cdot)$ and $\pi_i^R(\cdot)$ are independent of $T_i^*$ for any given $Z$.

By the first-order condition for $F_i$, $\lambda$ is equal to $|C|$, the measure of $C$. By the FOC for $p_i$,

$$\int_{z \in Z} a(z)\Big[q(p_i) + (p_i - c)q'\Big]dz - |C|a^m q(p_i) = 0, \tag{3.28}$$

so we have $p_i^* = c$ if and only if

$$\int_{z \in Z} a(z)q(p_i)dz = |C|a^m q(p_i), \tag{3.29}$$

or

$$\frac{1}{|C|}\int_{z \in Z} a(z)dz = a^m.$$

That is, marginal-cost pricing is optimal if (and only if) the share of calls going to network $A$ is the same for the average and the marginal consumer. An analogous argument applies to $\hat{p}_i$.

Quasiconcavity follows from (3.28) and (3.29): The objective function is increasing if $p_i < c$, flat at $p_i = c$, and decreasing otherwise. ■

**Proof of Lemma 9**

First, suppose there is a Nash equilibrium with *negative* externalities on both networks and a market share for network $A$ of $z \in (1-\mu, \mu)$ (this is more general than a symmetric outcome with $z = 1/2$). Then each mobile consumer on each network makes a share of $\gamma z + (1-\gamma)z = z$ of his calls to network $A$ and the complementary share to $B$, so Lemma 8 applies. In particular, profit increases as $p$ approaches $c$ and $\hat{p}$ approaches $\hat{c}$ as long as the set of actual consumers remains constant. This can be achieved by a corresponding change in the fixed fee $F$ as long as both firms create nonpositive externalities, so that market shares are determined as in (3.3) and (3.4). Given strictly negative externalities on the other network $j$, i. e. $p_j > \hat{p}_j$, firm $i$ can increase its payoff by a unilateral deviation to $p_i = \hat{p}_i$, eliminating the negative externalities on its network.

Now consider a symmetric equilibrium with $p_i = \hat{p}_i = p$ set by both firms $i = A, B$ for the mobile consumers, who are therefore indifferent between the two networks. Since $\hat{p}_A = \hat{p}_B$, access payments cancel out. Without loss of generality, suppose firm $A$ is chosen by no more than half of all mobile consumers. Firm $A$ can deviate to a tariff with $p_A < p < \hat{p}_A$ such that the members of group 1 (weakly) prefer $A$, and the members of group 2 choose $B$. That is, consumers choose networks as in Figure 3.2, and firm $A$ has at least as many customers as before. Moreover, each of $A$'s customers now makes a higher share of his calls on-net (so the average cost of each call is reduced), and $\hat{p}_A > \hat{p}_B$, resulting in a positive access balance in favour of $A$. Thus, the deviation to $p_A < \hat{p}_A$ (positive externalities) is profitable. ■

**Proof of Lemma 11**

Without loss of generality, let firm $A$ cede its mobile group-1 clientele to firm $B$, resulting in a profit of

$$\begin{aligned}\tilde{\pi}_A =& \frac{1-\mu}{2}\Big[\gamma(1-\mu)v(c) + (\gamma\mu + 1 - \gamma)v(\hat{c}) - f\Big] \\ &+ \frac{1-\gamma}{2}(1-\mu)mcq(\hat{c}) \\ &+ \frac{\gamma\mu}{2}(1-\mu)mcq(\hat{c}),\end{aligned}$$

where the first line is the profit from $A$'s captive consumers, the second line is the profit on incoming calls from group 2, and the last line is the profit on incoming calls from the mobile group 1 consumers who have switched to network $B$. Subtracting the shared-market profit (3.10) from $\tilde{pi}_A$ and collecting terms, the change in $A$'s payoff is given by

$$\tilde{\pi}_A - \pi^{\oplus} = \frac{\mu}{2}\Big[(2\gamma - 1 - \gamma\mu)(mcq(\hat{c}) + v(\hat{c}) - v) + f - v\Big]$$

This is a function of $m$ (since $\hat{c} \equiv (1+m)c$) whose value at $m = 0$, $(f - v(c))\mu/2$, is negative by Assumption 4. Moreover, the first derivative w. r. t. $m$,

$$\frac{\mu}{2}(2\gamma - 1 - \gamma\mu)mc^2q'(\hat{c}),$$

is non-positive for all $m \geq 0$. ■

**Proof of Proposition 4**

Consider the symmetric candidate equilibrium with positive externalities on both networks and tariffs $T_i = (F_i, c, \hat{c})$, where $F_i = \gamma v(c) + (1-\gamma)v(\hat{c})$ by Lemma 10, $i = A, B$. Suppose firm $A$ deviates and attracts $B$'s mobile consumers set leaving usage prices constant at marginal cost. Then it must adjust $F_A$ to satisfy the incentive constraint of the mobile consumers in group 2,

$$F_A^D = (1 - \gamma + \gamma\mu)v(c) + (1-\mu)\gamma v(\hat{c}).$$

Plugging $F_A^D$, $p_A^D = c$ and $\hat{p}_A^D = \hat{c}$ into (3.11), one obtains the profit from this suboptimal deviation for $A$,

$$\begin{aligned}&\frac{1-\mu}{2}\Big[(\gamma + (1-\gamma)\mu)v(c) + (1-\gamma)(1-\mu)v(\hat{c}) - f\Big] \\ &+ \mu(F_A^D - f) + \frac{1-\mu}{2}(1 - \gamma + \gamma\mu)mcq(\hat{c}).\end{aligned}$$

Comparing this with the shared-market payoff in (3.10), one can see that deviating pays if

$$v(c) - f \geq (4\gamma - 2 - 3\gamma\mu + \mu)\big(v(c) - v(\hat{c})\big) - (2\gamma - 1 - \gamma\mu)mcq(\hat{c}).$$

But $v(c) > f$ by Assumption 4, and the RHS is zero for $m = 0$. By continuity, there is a value $\underline{m} > 0$ such that the inequality holds for all $m \in [0, \underline{m})$. ■

**Proof of Lemma 12**

Without loss of generality, suppose that firm $A$ deviates from the shared-market outcome and attracts both sets of mobile consumers with a tariff $T_A^D$. For all mobile consumers to subscribe to $A$, neither those from group 1 nor those from group 2 must have an incentive to defect. Moreover, the members of both groups must not prefer to coordinate on defecting together. This implies the following incentive constraints. (Since by Lemmas 8 and 10, $B$'s tariff has $p_B = c$, $\hat{p}_B = \hat{c}$ and $F_B = \gamma v(c) + (1-\gamma)v(\hat{c})$.)

First, the mobile members of group 1 must prefer $A$ to $B$ given that the mobile group 2 members also choose $A$:

$$\begin{aligned} F_A^D \leq\ & (\gamma + \mu - \gamma\mu)v(p_A^D) + (1-\gamma)(1-\mu)v(\hat{p}_A^D) \\ & + (2\gamma - 1)(1-\mu)\big(v(c) - v(\hat{c})\big). \end{aligned} \tag{3.30}$$

Second, the group-2 consumers must choose $A$ given that so does group 1:

$$F_A^D \leq (1 - \gamma + \gamma\mu)v(p_A^D) + (1-\mu)\gamma v(p_A^D). \tag{3.31}$$

Third, a joint defection to $B$ of all mobile consumers would leave the mobile members of group 1 worse off if

$$F_A^D \leq \gamma v(p_A^D) + (1-\gamma)v(\hat{p}_A^D) + (2\gamma - 1 - \gamma\mu)\big(v(c) - v(\hat{c})\big). \tag{3.32}$$

Clearly, when

$$v(p_A^D) - v(\hat{p}_A^D) \geq -\big(v(c) - v(\hat{c})\big), \tag{3.33}$$

then (3.31) implies both (3.30) and (3.32). Assuming that (3.33) holds, the objective is to maximize (3.11) subject to (3.31). This is equivalent to the simplified program

$$\begin{aligned} & \max_{T_A^D} (1+\mu)(p_A^D - c)q(p_A^D) + (1-\mu)(\hat{p}_A^D - \hat{c})q(\hat{p}_A^D) + 2F_A^D \\ & \quad \text{s.t. } F_A^D \leq (1 - \gamma + \gamma\mu)v(p_A^D) + (1-\mu)\gamma v(p_A^D). \end{aligned} \tag{3.34}$$

The tariff given in the Lemma follows from the first-order conditions. ■

**Proof of Lemma 13**

To prove $(i)$, it needs to be shown that

$$v(p^D) - v(\hat{p}^D) \geq -\Big(v(c) - v(\hat{c})\Big) \tag{3.35}$$

for all $m$ larger than some $\overline{m} > 0$, with $p^D$ and $\hat{p}^D$ from Lemma 12. For $m = 0$, $p^D > c$ and $\hat{p}^D < c$, so the LHS of the above inequality is negative, and the RHS is zero. By continuity, the inequality fails for all $m$ in an open interval around zero. As $m$ increases, the RHS decreases monotonically at a rate of $cq(\hat{c})$, while $p^D$ is constant in $m$. Hence, it suffices to show that $v(\hat{p}^D)$ decreases in $m$ and eventually becomes smaller than $v(p^D)$.

By Lemma 12,

$$\hat{p}^D = (1+m)c + (2\gamma - 1)\frac{q(\hat{p}^D)}{q'(\hat{p}^D)}.$$

Interpret the RHS as a function of $p$ (with parameter $m > 0$) and observe that

$$\frac{\partial}{\partial p}\frac{q(p)}{q'(p)} < 1 \quad \Leftrightarrow \quad q(p)q''(p) > 0,$$

so the RHS has a unique fixed point in $p$, which defines $\hat{p}^D$. An increase in $m$ unambiguously shifts the RHS upwards and $\hat{p}^D$ to the right. Thus, the LHS of (3.35) increases monotonically in $m$. Moreover, the derivative of the RHS w. r. t. $m$ is larger than $-\infty$, so $\hat{p}^D$ goes out of bounds as $m \to +\infty$, and $v(p^D) - v(\hat{p}^D)$ eventually becomes positive.

As for $(ii)$, $\pi^D(T^D)$ is the true deviation profit as long as (3.35) holds, i. e. when $m \geq \overline{m}$. When $m < \overline{m}$, it can be shown the the incentive constraints for the mobile consumers in *both* groups must be binding when firm $A$ optimally deviates from the shared-market equilibrium. (Assume that only the constraint for the group-1 users is binding while the constraint for the mobile consumers in group 2 is slack. Then the optimal deviation tariff satisfies (3.35), so the constraint for the group-2 users is necessary and sufficient, a contradiction.) This amounts to an additional constraint for the program in (3.34), so the resulting deviation profit can be no larger than $\pi^D(T^D)$. ■

**Proof of Lemma 14**

Suppose $\gamma = 1$ and $m \geq \overline{m}$. By (3.12) and (3.14), $f_{\min} < f^{\max}$ iff

$$(1+\mu)(p^D - c)q(p^D) + (1-\mu)(\hat{p}^D - \hat{c})q(\hat{p}^D) \\ + 2F^D + (1-\mu)mcq(\hat{c}) - v(c) \quad < \quad v(c).$$

Since $\hat{c} = c + mc$, this is equivalent to

$$(1+\mu)(p^D - c)q(p^D) + (1-\mu)(\hat{p}^D - c)q(\hat{p}^D) \\ + 2F^D + (1-\mu)mc\Big[q(\hat{c}) - q(\hat{p}^D)\Big] \quad < \quad 2v(c).$$

Since $\hat{p}^D < \hat{c}$ by Lemma 12, the term in square brackets on the LHS is negative. Hence, it suffices to show that

$$(1+\mu)(p^D - c)q(p^D) + (1-\mu)(\hat{p}^D - c)q(\hat{p}^D) + 2F^D \leq 2v(c). \tag{3.36}$$

Suppose that the LHS of (3.36) was maximized subject to the participation constraint of the *average* consumer,

$$F^D \leq \frac{1+\mu}{2}v(p^D) + \frac{1-\mu}{2}v(\hat{p}^D).$$

Then the optimal tariff $T^D$ would consist of $p^D = \hat{p}^D = c$ and $F^D = v(c)$, so that the LHS of (3.36) would just equal $2v(c)$. When the average and marginal consumer differ, as they do for the firm deviating from the shared-market outcome, then some rent is left to the inframarginal consumers, and the inequality in (3.36) is strict. By continuity, there is a $\underline{\gamma} < 1$ such that this result holds for all $\gamma \in (\underline{\gamma}, 1]$. ■

**Proof of Lemma 15**

Suppose instead that there was a symmetric equilibrium where both firms set $p < \hat{p}$ given $m < 0$. By Lemma 7, all the mobile members of group choose the same network. Moreover, the only symmetric outcome is the one where the members of group 1 all choose firm $A$ and the consumers in group 2 join network $B$. Hence, the mobile clientele of each firm is homogeneous, and Lemma 8 applies. In particular, profits are quasiconcave in usage prices (holding market shares constant), with a maximum at $p = c$ and $\hat{p} = \hat{c}$.

The incentive constraint for the group-1 consumers to stay on network $A$ is

$$F_A \leq F_B + \gamma v(p_A) + (1-\gamma)v(\hat{p}_A) - (1-\gamma+\gamma\mu)v(p_B) - (1-\mu)\gamma v(\hat{p}_B).$$

For the mobile members of group 2 to remain with firm $B$, we need

$$F_A \geq F_B + (1-\gamma)v(p_A) + \gamma v(\hat{p}_A) - \gamma v(p_B) - (1-\gamma)v(\hat{p}_B).$$

The two constraints are consistent (there is a non-empty set of values for $F_A$ such that both are satisfied) if

$$(2\gamma - 1 - \gamma\mu)(v(p_B) - v(\hat{p}_B)) \geq (2\gamma - 1)(v(\hat{p}_A) - v(p_A)). \tag{3.37}$$

By Lemma 8, the profit of firm $A$ increases monotonically as its usage prices approach $p = c$ and $\hat{p} = \hat{c} < c$. Since $2\gamma - 1 - \gamma\mu > 0$ by Assumption 6, and $v(p_B) > v(\hat{p}_B)$ when firm $B$ sets $p_B < \hat{p}_B$, the LHS of (3.37) is strictly positive, firm $A$ can approach marginal-cost pricing to a point where $\hat{p}_A < p_A$ and still satisfy the inequality. Hence, whenever one of the firms creates positive externalities, the best reply of its rival, holding market shares constant, is to set $\hat{p} < p$, so there is no shared-market equilibrium with positive externalities when $m < 0$. ■

**Proof of Lemma 16**

Since $\hat{c} \equiv (1+m)c$, $g(0) = 0$. For $m < 0$,

$$\frac{\partial g}{\partial m} = -ycq(\hat{c}) + mc^2 q'(\hat{c})$$

is positive throughout if and only if $y = 0$. In this case, $g$ is an increasing function of $m$ with a maximum of zero at $m = 0$. If $y > 0$, however, $g$ decreases toward zero for small (absolute) values of $m$: As $m \to 0$, $\partial g/\partial m$ converges to $-ycq(c) < 0$ ($q'(c)$ is finite $\forall c > 0$). Thus, when $y > 0$, then $g$ is positive and approaches zero from above for all values of $m$ in a non-empty interval $(\tilde{m}, 0)$. ■

**Proof of Lemma 17**

Given $F^{\ominus}$ from (3.21), the participation constraint in (3.17) can be rewritten as

$$v(c) - f > \frac{\mu}{2}\big(v(\hat{c}) - v(c)\big),$$

where the LHS is strictly positive and constant in $m$, and the RHS is zero for $m = 0$ and monotonically decreasing in $m$. Thus, the inequality holds for $m \approx 0$ and changes sign at most once, at $m^{PC} < 0$. (Note that $m^{PC}$ need not be finite.) If $m < m^{PC}$, then $F^{\ominus} = \big(v(\hat{c}) + v(c)\big)/2$, and the derivative of the shared-market profit in (3.22) w. r. t. $m$ becomes

$$\frac{\partial \pi^{\ominus}}{\partial m} = \frac{mc^2 q'(\hat{c})}{4} \qquad > 0 \; \forall m < 0.$$

■

**Proof of Lemma 19**

Without loss of generality, suppose firm $A$ corners the mobile market. Let $p_B^M$ denote the usage price offered to the mobile consumers by firm $B$. (In equilibrium, this offer is not taken up by any consumer; $p_B^M$ only affects payoffs off the equilibrium.) The proof consists of two steps. First, it is shown that the incentive constraints for the two firms together imply a lower bound for $\left|p_B^M - c\right|$. Second, any $p_B^M$ satisfying this constraint is shown to be weakly dominated. For simplicity, the attention is restricted to $m > 0$; a completely analogous argument applies when $m < 0$.

$(i)$ *Lower bound for* $p_B^M$. Consider the incentive for firm $A$ to deviate from the candidate equilibrium. Precisely, let $A$ cede the mobile market to firm $B$ given that $B$ has set a tariff $T_B^{LR} = (F^{LR}, p^{LR})$ for its captive clients, and $T_B^{MR} = (F^{MR}, p^{MR})$ for the mobile consumers. This deviation is profitable for $A$ unless

$$A \geq B + C + \frac{1-\mu}{2} mc\mu\big[q(p^{MR}) - q(p^{LR})\big],$$

where

$$\begin{aligned}
A &\equiv \mu\Big((p^{MK} - c)q(p^{MK}) + F^{MK} - f\Big) \\
B &\equiv \frac{1-\mu}{2}\Big((p^{LR} - c) + v(p^{LR}) - (p^{LK} - c)q(p^{LK}) - v(p^{LK})\Big) \\
C &\equiv \frac{1-\mu}{2} mc\Big[(1-\gamma)(1-\mu)q(p^{LK}) + \mu q(p^{MK}) - (1-\gamma+\gamma\mu)q(p^{LR})\Big]
\end{aligned}$$

On the other hand, firm $B$ can deviate from the candidate outcome by capturing the mobile market. Given usage prices $p_A^{LK}$ and $p_A^{MK}$ set by firm $A$ for its locked-in and mobile customers, respectively, this deviation is profitable for firm

$B$ unless

$$A \leq B + C + \frac{1-\mu}{2} mc\mu \big[q(p^{MK}) - q(p^{LK})\big].$$

Thus, for the two incentive constraints to be simultaneously satisfied, it must be that

$$m\Big[q(p^{MK}) - q(p^{MR}) + q(p^{LR}) - q(p^{LK})\Big] \geq 0.$$

Since $p^{LK} \leq \min\Big\{p^{MK}, p^{LR}\Big\}$ for all positive values of $m$, this implies that

$$p^{MR} \geq \max\Big\{p^{MK}, p^{LR}\Big\} \;\forall m > 0. \tag{3.38}$$

*(ii) Weak-dominance argument.* Fix a tariff $T_B^{MR} = (F_B^{MR}, p_B^{MR})$ which satisfies (3.38) and consider the action space of firm $A$ in the mobile market. As long as

$$v(p_A^M) - F_A^M > v(p_B^{MR}) - F_B^{MR}, \tag{3.39}$$

firm $B$ is indifferent about $T_B^{MR}$ because its offer is not taken up by any consumers. If (3.39) is reversed, then firm $B$'s payoff increases as $p_B^{MR}$ approaches $p^{MK}$ (because $\pi_i^K$ from (3.25) is quasi-concave in $p_i^M$). In the remaining case where consumers are indifferent, $B$'s profit is quasi-concave in $p^{MR}$ with a maximum at $p^{MS}$ (see (3.24)). Consequently, a usage price of $\bar{p}^{MR} \equiv \max\{p^{MK}, p^{MS}\}$ weakly dominates any $p^{MR} > \bar{p}^{MR}$. But $p^{MS} = c + mc/2$ is less than $p^{LR} = c + (1 - \gamma + \gamma\mu)mc$ whenever $\gamma \leq 1/(2 - 2\mu)$, for which $\mu > 1/2$ is sufficient. Thus, $\bar{p}^{MR} \leq \max\{p^{MK}, p^{LR}\}$, and any $p^{MR}$ satisfying (3.38) is weakly dominated by $\bar{p}^{MR}$. ■

# Chapter 4

# Bargaining, Mergers, and Heterogeneous Outsiders

With Konrad Stahl

## 4.1 Introduction

Despite the prevelance in economic theory of markets with many players on one or both sides, trade in real life very often takes place on a bilateral basis with a single buyer and seller bargaining over prices and quantities. This is especially true for intermediate goods traded within the kind of multi-tiered, oligopolistic market structure often referred to as an "industry". Of particular interest in such a setting is the relationship between market structure and bargaining outcomes.

This is more than just a theoretical exercise. Over the past two decades, mergers and acquisitions have played a very prominent role in the evolution of many important industries.[1] DaimlerChrysler, AOL Time Warner and Novartis are just three of the most well-known examples; others like the GE–Honeywell merger failed to become reality only because of exogenous obstacles. Somewhat less spectacular but nonetheless important for their industries, there have been a number of de-mergers like the spin-off of Delphi from General Motors. Yet the economic theory of mergers remains very incomplete. The explanations of merger incentives range from monopolization motives to "synergies" and empire-building by ill-controlled managers. Another aspect which has attracted some attention is the impact of a merger on bargaining shares.[2] However, not very much is

[1]See, for instance, Gugler et al. (2001).
[2]Segal (2000) gives an overview of the literature on integration and bargaining.

known about the bargaining effects of mergers in multi-tiered industries with heterogeneous firms; nor is there yet a fully convincing theoretical explanation of merger waves.

This paper is a contribution towards the development of a more general theory of bargaining and mergers in industries with bilateral trade among very different players. We build on the cooperative approach recently developed by Segal (2000), paying particular attention to the role of heterogeneity among the outsiders to a merger. Abstracting from allocative effects and concentrating on bargaining considerations alone, we find that heterogeneity among outsiders gives rise to a number of new effects: Outsiders can be the main beneficiaries of a profitable merger; the final consequences of player-specific shocks can be quite different from the initial effect; and the merger decisions of two disjoint pairs of firms are generally strategic complements or substitutes, suggesting a bargaining-based rationale for mergers to occur in waves.

Throughout, we take the candidate players for a merger as exogenously given. In this assumption, and in the flavour of some of our results, this paper is related to the exogenous-mergers literature following Stigler (1950), which focuses on mergers among homogeneous firms in a single market. A standard reference is Salant et al. (1983), who show that the merging firms in a Cournot oligopoly with constant marginal cost must have a market share of at least 80 percent in order for the merger to be privately profitable. Heuristically, the merging firms move the industry allocation closer to the cartel outcome (by internalizing the external effects they exert on each other) but lose some market power by becoming a single player instead of several independent ones; the net effect is ambiguous for the merging firms but positive for all outsiders. More recently, Nilssen and Sørgard (1998) have extended this approach to study the interaction between disjoint mergers within the framework of Fudenberg and Tirole (1984). However, the combination of allocative and market-power effects, as well as the marginal-cost synergies assumed by Nilssen and Sørgard, make it somewhat difficult to consider each effect in isolation even in a simple, horizontal market.

In a related strand of the literature, the interaction of upstream and downstream players is taken into consideration in the analysis of merger incentives within a non-cooperative framework. Examples include the application to union formation in Horn and Wolinsky (1988b) and Stole and Zwiebel (1996b); Stole and Zwiebel (1998); and most recently the supplier–retailer model of Inderst and Wey (2001). A notable improvement of the last paper over some predecessors like Horn and Wolinsky (1988a) lies in the use of "efficient" bargaining (i. e. over quantities as well as prices), which eliminates any allocative effects of a merger and allows for an analysis based entirely on bargaining-power considerations. Broadly speaking, a merger in this setting—like in Horn and Wolinsky (1988b)—is profitable if the merging players face a common supplier operating at increasing unit cost, or a common customer who regards their output as substitutable. In either case, a merger shifts the focus of each bilateral bargaining session towards the higher inframarginal rents, resulting in an increased payoff for the merged entity.

Segal (2000) takes the analysis of mergers ("collusion", in his terms) to another level using a cooperative approach in which the joint profit of any coalition of firms is always maximized by definition, and where the game is entirely about the distribution of rents.[3] In contrast with the very specific, non-cooperative models in the earlier literature, Segal's approach leads to a unifying view of the underlying bargaining effects in a very general setting. Among other things, he shows that the profitability of a merger is determined by the *impact* of the outside players on the substitutability of the merging firms rather than the *level* of that substitutability per se.

A limitation of almost all previous work, however, lies in the way outsiders are modeled. Since the focus is generally on merger incentives as such, relatively little attention is paid to outsiders' characteristics. To simplify matters, they are usually taken to be essentially homogeneous. For instance, while the basic model in Inderst and Wey (2001) allows for heterogeneity, no attention is given to the case where one supplier has increasing unit costs while the other operates under economies of scale. Segal (2000) employs a robustness condition with respect to the choice of random-order values which implies that all outsiders must either increase or decrease the merging players' substitutability, ruling out the interesting and realistic case where there are both types of outsiders.

While we follow Segal's very general cooperative approach, we concentrate on the effects of a merger on heterogeneous outsiders, restricting our attention to Shapley values rather than general random-order values. Relative to Segal, this implies that we take intrinsic bargaining abilities as known ex ante. Regardless of this restriction, individual payoffs depend on the marginal contributions of each player to various preceding coalitions; heuristically, these correspond to different

[3]Inderst and Wey (2001) obtain an equivalent setting by a combination of efficient bargaining and assumptions ruling out coordination failure.

bargaining situations where some of the other players refuse to cooperate.

Mergers or de-mergers in our model affect payoffs by changing the composition of the preceding coalitions in each ordering of the players. Intuitively, this corresponds to a change in the set of different bargaining situations in which each player finds himself after some of the others have refused to cooperate. A recurring theme will be the complementarity of two merging players in the presence of some outsiders. When two players have merged, they employ or withdraw their joint resources together rather than separately. This has an impact on the average marginal contributions of the other players because it replaces each bargaining situation where only one of the merging parties is active by one in which either both or neither of them participate. If the merging parties are complementary in their impact on the remaining players, then their joint impact exceeds the sum of their individual effects. We make this argument more precise below using a second-difference operator introduced by Segal (2000) (and earlier, by Ichiishi (1993)) to capture the complementarity of the merging players.

A key feature of our analysis is that we allow outsiders to be heterogeneous. In real life, bargaining takes place in fairly complex settings, where the players can face very different bilateral-trading partners at the same time. For example, producers of intermediate inputs deal with both downstream customers and upstream suppliers, and there is no reason to assume that these two kinds of trading partners share any fundamental characteristics. The same is true in the case of a vertical merger. Even the outsiders to a horizontal merger in a simple two-sided market can very easily be fundamentally different. For example, there is nothing exceptional about a firm dealing simultaneously with some suppliers producing at increasing unit cost, and with others enjoying economies of scale.

The automobile industry is a case in point. Suppliers to car makers range from small firms providing specialized springs exclusively to the industry at increasing average cost, to large steel producers supplying raw metal sheets to many industries at strongly decreasing average cost. On their output side, some car makers deal with very different retailing channels. For instance, car brands or model types might be substitutes in small local markets but complements for a large multi-brand retailer trying to expand its geographical reach.

Outsider heterogeneity is not only an important empirical phenomenon; it also leads to effects which are otherwise absent in pure bargaining models. If all outsiders are fundamentally alike, then a merger in a zero-sum game can only be profitable if it hurts all outsiders. It follows immediately that the heterogeneity of outsiders is a necessary condition for some of them to benefit from a profitable merger. Indeed, when outsiders differ sufficiently, then the main effect of a merger can be a redistribution of surplus from some outsiders to others, while the merging firms themselves receive only a small share of the redistributed rents.

This is of particular interest when an outsider is hit by an exogenous shock prior to the merger decision. Even a shock specific to that outsider can have an indirect effect, in addition to the obvious direct one. This is because the characteristics of the affected outsider have an influence on the contributions of other players and, therefore, on how they are affected by a merger. By changing the amount of surplus received by each outsider after a merger, the shock can change the merger decision itself, which in turn can more than compensate the payoff effect of the shock originally experienced by the affected outsider.

This has important implications for empirical studies. First of all, the correlation between firm-specific shocks and profits is distorted or even reversed when these shocks result in mergers (or de-mergers) in the same industry. (Note that an "industry" in our context includes all the firms which are part of the same buyer-supplier network, so the merger need not occur in the immediate vicinity of the firm affected by the shock.) Second, the main incidence of the shock may fall on players who are only indirectly connected to the one originally affected. Re-interpreting the exogenous shock as the (binding) technology adoption of an outsider, we also conclude that the prospect of a merger can lead to inefficient incentives for technology choice and R&D effort.

The existence of different kinds of outsiders is also a necessary and generically sufficient condition for the interaction of two disjoint mergers. When there are additional outsiders who do not take part in either merger, then the two mergers interact symmetrically, turning merger decisions into strategic substitutes or complements. This qualifies an intermediate result derived by Inderst and Wey (2001), who find that the merger decisions of two disjoint pairs of players are independent of each other in a four-firm industry. Our results show that this does not generalize to a setting with more than four firms because each merger affects the impact of the other on the payoffs of those players who participate in neither one.

The remainder of this paper is organized as follows. The basic model is introduced in the next section, along with a measure of merger effects based on the Shapley value. We motivate our use of a cooperative model in Section 4.7.1 in the Appendix, where we show that the Shapley value is a useful solution concept in a very general setting of bilateral bargaining if one abstracts from allocative externalities and coordination problems to concentrate exclusively on bargaining effects.[4] In spite of some limitations of our non-cooperative foundation (for instance, there is no extensive form for the simultaneous bilateral negotiations), we take this as an argument for following and extending Segal's general cooperative approach as against specific non-cooperative models which are subject to similar limitations without offering the same generality of results. Our notion of heterogeneity is defined in Section 4.3 and shown to be necessary for outsiders to benefit from a profitable merger. Section 4.4 contains the analysis of an exogenous shock experienced by one of the outsiders before the merger decision. In Section 4.5 we consider the interaction between two mergers. Our main results are illustrated in two examples in Sections 4.4.1 and 4.5.4. Section 4.6 concludes. Some proofs are contained in Section 4.7.2 in the Appendix.

## 4.2 The model

Consider a cooperative game with transferable utility among a set of players $N = \{1, \ldots, n\}$. For instance, $N$ can be thought of as an industry consisting of $n$ firms facing each other in various vertical and horizontal bilateral relationships. As usual, the characteristic function $v : 2^N \to \mathbb{R}_+$ defines the worth of any coalition $S \subseteq N$. This is the maximum total net profit which the members of $S$ can achieve collectively if they coordinate their actions in the absence of the remaining players $N \setminus S$. The worth of the empty set is normalized to zero, $v(\emptyset) = 0$. Note that $v(S)$ depends only on the resources available to the players in $S$; the allocation of control rights over those resources affects individual payoffs but not the overall worth of the coalition $S$.

A merger of two players $a$ and $b$ in $N$ is the binding decision to face all other players as a single entity which simultaneously contributes (or withdraws) the joint resources of its constituent parts in any given bargaining situation. Likewise, a de-merger is the decision to separate an entity into two independent players each of whom controls some of the joint resources. We assume that there is at most one technologically feasible way of partitioning the resources of any player into two independent units. Moreover, neither a merger nor a demerger affects the joint contribution of both players to any given coalition in $N$; that is,

[4]In contrast with Inderst and Wey (2001), we motivate the Shapley value by a stability concept generalized from Stole and Zwiebel (1996), rather than by contingent contracts.

we explicitly rule out synergies and diseconomies of scale.

At some points we will assume that $v(S)$ is a continuous function of a parameter $\lambda$. What we have in mind is a parameter with a continuous effect on the underlying profit functions of one or several of the players in $N$ and, hence, on the optimal joint payoff which any given coalition $S$ can achieve. For example, $\lambda$ might represent the marginal cost of a supplier or the number of final consumers.

The game is solved by the Shapley (1953) value. Let $\Pi$ denote the set of all $n!$ permutations of the $n$ players, and let $\pi \in \Pi$ denote a particular permutation (or ordering). Further, let

$$\Delta_i(S) \equiv v(S \cup i) - v(S)$$

denote the marginal contribution of player $i \notin S$ to coalition $S \subset N$.[5] As is well known, the Shapley value is derived axiomatically and prescribes that each player receive his average marginal contribution to the coalition of players preceding him in each ordering $\pi \in \Pi$, with the same probability weight of $1/n!$ accorded to each ordering.

Let $\prec_\pi$ denote precedence in ordering $\pi$, so that

$$\{a, b\} \prec_\pi \{c, d\}$$

if and only if players $a$ and $b$ precede players $c$ and $d$ in ordering $\pi$, regardless of the pairwise orderings of $a$ and $b$, and $c$ and $d$ respectively. Let

$$B(i, \pi) \equiv \{k \in N : k \prec_\pi i\}$$

denote the set of players preceding player $i$ in ordering $\pi$. In other words, $B(i, \pi)$ is the coalition of players to which $i$ makes his marginal contribution in this ordering.

The Shapley value for player $i$ is given by

$$Sh^i = \frac{1}{n!} \sum_{\pi \in \Pi} \Delta_i(B(i, \pi)).$$

Non-cooperative foundations for the Shapley value have been given by Gul (1989), Hart and Mas-Colell (1996), and Stole and Zwiebel (1996a) among others.

[5] For simplicity, we denote the single-element set $\{i\}$ as $i$. The difference-operator notation is borrowed and adapted from Segal (2000), who credits Ichiishi (1993) with an earlier introduction of the same concept to describe complementarities between players in cooperative games.

To motivate our use of the Shapley value, Section 4.7.1 in the Appendix contains a generalization of the stability condition developed by Stole and Zwiebel (1996a). At the core of this concept is the assumption that renegotiation is always possible. Thus, disagreement payoffs in any bilateral bargaining situation are given by what the players can achieve after a general renegotiation of all active contracts rather than by what is specified in the equilibrium set of contracts.[6] We present a setup where the firms in an industry are connected through a set of bilateral trading links. The terms of trade are also determined bilaterally by delegated, simultaneous, and efficient bargaining. Stole–Zwiebel stability in this setup implies that payoffs obey Myerson's (1980) "fair allocation" rule, and hence, by his central result, are equal to the players' Shapley values.

### 4.2.1 A measure of merger effects

This section is based largely on Segal (2000), who analyzes mergers as a combination of exclusive and inclusive contracts, arriving at essentially the same measure of merger effects (4.4) in a closely related way.[7]

The impact $M^k(ab)$ of a merger of any two firms $a$, $b$ in $N$ on an outsider $k \in N \setminus \{a, b\}$ is the difference between her Shapley values after and before the merger,

$$M^k(ab) \equiv Sh^k\big|_{a \text{ and } b \text{ merged}} - Sh^k\big|_{a \text{ and } b \text{ separate}},$$

where the Shapley value after the merger of $a$ and $b$ is evaluated over all permutations of the remaining $n-1$ players; one of these is the merged firm whose marginal contribution to a preceding coalition $B$ is the sum of the contributions of its components $a$ and $b$, $\Delta_a(B) + \Delta_b(B \cup a)$. To determine $M^k(ab)$, it is useful to introduce the notion of a dummy player.

**Definition 1** *A player $i \in N$ is called a dummy player if and only if $v(S \cup i) = v(S \setminus i)$ for all $S \subseteq N$.*

Using this definition, we can evaluate post-merger Shapley values over the permutations of the original $n$ players, which allows us to compare the effect of the merger for each permutation.

[6]Inderst and Wey (2001) achieve the same result by letting firm bargain over sets of contracts contingent on the outcomes of all other bilateral negotiations.

[7]Segal also allows for more general random-order values, assuming only that the probability distribution over all orderings is symmetric with respect to the merging players; this is trivially satisfied by the Shapley value.

It is easily checked that adding a single dummy player to the set of $n-1$ post-merger players has no effect on the Shapley values of the $n-1$ "real" players. Let $N_m$ denote the set of players consisting of the merged firm, the $n-2$ outsiders in $N \setminus \{a, b\}$, and the dummy player. Let $\Pi_m$ be the set of $n!$ permutations of the players in $N_m$. The post-merger payoff to an outsider $k \in N \setminus \{a, b\}$ is equal to

$$Sh^k\big|_{a \text{ and } b \text{ merged}} = \frac{1}{n!} \sum_{\pi_m \in \Pi_m} \Delta_k(B(k, \pi_m)).$$

For each ordering $\pi \in \Pi$ there is exactly one counterpart $\pi_m \in \Pi_m$ where the merged firm takes the position of $a$ in $\pi$ and the dummy player appears instead of $b$. Hence, to calculate $M^k(ab)$ we need only determine the difference in the set of players preceding $k$ in each $\pi$ and its corresponding $\pi_m$.

Clearly, when both $a$ and $b$ precede $k$ in $\pi$, then the coalition preceding $k$ in $\pi_m$ differs only in the presence of the dummy player, so that the marginal contribution of $k$ is unchanged. Similarly, there is no effect when both $a$ and $b$ follow $k$ in $\pi$. The only relevant orderings are those where $k$ comes between $a$ and $b$.

More formally, let

$$B'(k, \pi) \equiv B(k, \pi) \setminus \{a, b\}$$

denote the set of players preceding $k$ except for the merging firms $a$ and $b$. In an ordering $\pi \in \Pi$ with $a \prec_\pi k \prec_\pi b$ (where $k$ follows $a$ but precedes $b$), player $k$ makes her marginal contribution to $B'(k, \pi) \cup a$. In the corresponding post-merger ordering $\pi_m$, $k$ makes her marginal contribution to $B'(k, \pi) \cup \{a, b\}$, and the net effect on her payoff is equal to $\Delta_k(B'(k, \pi) \cup \{a, b\}) - \Delta_k(B'(k, \pi) \cup a)$.

On the other hand, when $b \prec_\pi k \prec_\pi a$, then player $k$'s contribution in $\pi_m$ is effectively made to $B'(k, \pi)$ (plus the dummy player), resulting in a payoff difference of $\Delta_k(B'(k, \pi)) - \Delta_k(B'(k, \pi) \cup b)$. With some additional notation, these payoff effects can be summed up quite concisely over all the relevant orderings.

Let

$$\begin{aligned} \Delta^2_{ij}(S) &\equiv \Delta_j(\Delta_i(S)) \\ &\equiv \Delta_i(S \cup j) - \Delta_i(S) \\ &\equiv v(S \cup \{i, j\}) + v(S) - v(S \cup i) - v(S \cup j) \end{aligned} \tag{4.1}$$

denote the effect of player $j \in N$ on the marginal contribution of player $i \in N$ to a preceding coalition $S \subseteq N \setminus \{i, j\}$.[8] As the last line in (4.1) shows, this can be

[8]Note that the definition of the second-difference operator implies that $\Delta^2_{ij}(S) \equiv \Delta^2_{ji}(S)$ for $i \neq j$.

interpreted as a measure of the supermodularity of $v$ on $\{i, j\}$ or, equivalently, of the complementarity between $i$ and $j$ in the presence of the players in $S$.[9]

Using the definition in (4.1), the payoff differences for $k$ between orderings in $\Pi$ and $\Pi_m$ can be expressed as

$$\begin{cases} \Delta^2_{bk}(B'(k,\pi) \cup a) & \text{if } a \prec_\pi k \prec_\pi b, \\ -\Delta^2_{bk}(B'(k,\pi)) & \text{if } b \prec_\pi k \prec_\pi a, \\ 0 & \text{otherwise.} \end{cases} \tag{4.2}$$

Further, let the third-difference term

$$\begin{aligned} \Delta^3_{ijk}(S) &\equiv \Delta_k(\Delta^2_{ij}(S)) \\ &\equiv \Delta^2_{ij}(S \cup k) - \Delta^2_{ij}(S), \end{aligned} \tag{4.3}$$

denote the impact of player $k$ on the complementarity between players $i$ and $j$ in the presence of the players in $S$.[10] For each ordering $\pi \in \Pi$ with $a \prec_\pi k \prec_\pi b$ there is exactly one $\pi^* \in \Pi$ which is identical except for the transposition of $a$ and $b$ (and which, by the definition of the Shapley value, has the same probability weight as $\pi$). Hence, the terms in (4.2) can be pairwise combined and summed up over those orderings where $a \prec_\pi k \prec_\pi b$. Defining the binary indicator function

$$1[expr] = \begin{cases} 1 & \text{if } expr \text{ is true} \\ 0 & \text{otherwise,} \end{cases}$$

the effect of the merger of $a$ and $b$ on the payoff of $k$ can be expressed as

$$M^k(ab) = \frac{1}{n!} \sum_{\pi \in \Pi} 1[a \prec_\pi k \prec_\pi b] \Delta^3_{kab}(B'(k,\pi)). \tag{4.4}$$

We will refer to $M^k(ab)$ as the *merger benefit* for $k$, without implying that the benefit is necessarily positive.

The third-difference terms $\Delta^3_{kab}(B'(k,\pi))$ in (4.4) can be interpreted in two different ways. One is to see them as the impact of player $k$ on the complementarity of the merging parties $a$ and $b$ in the presence of the preceding outsiders $B'(k,\pi)$. The more player $k$ increases this complementarity, the more she benefits from the merger of $a$ and $b$. Heuristically, consider two sellers $a$ and $b$ and

[9]See Segal (2000) for a discussion of the difference operator $\Delta$.

[10]The term $\Delta^3_{ijk}(S)$ measures the impact of *any* of the three players $i, j, k$ on the complementarity between the remaining two. This is another implication of the fact that the difference operator is symmetric with respect to the players over which the differences are taken.

a single buyer $k$ bargaining over two goods, one delivered by $a$ and the other by $b$. In the absence of $k$, the profit of $a$ and $b$ is zero, as is their complementarity $\Delta^2_{ab}(\emptyset)$.[11] With $k$ in the game, $\Delta^2_{ab}(k) \equiv v(abk) + v(k) - v(ak) - v(bk)$ is positive if the two goods are complements and negative if they are substitutes.[12] Hence, the third-difference term $\Delta^3_{kab}(\emptyset)$ is positive if and only if player $k$ regards the goods as complements. As is well known, $k$ benefits from the merger of $a$ and $b$ in this case.[13]

However, while the complementarity of the traded goods coincides with that of the players $a$ and $b$ in the example just given, the two notions of product and player complementarity are strictly different concepts. Indeed, one of the main contributions of Segal (2000) is to demonstrate that merger incentives depend on changes in the complementarities among players, not goods, and that some of the paradoxes reported in earlier research are due to a confusion of the two concepts. For illustration, add a fourth player $u$ in our simple 3-player economy and suppose that $u$ produces a common input to $a$ and $b$ at increasing unit cost. If $u$'s unit cost increase is sufficiently strong relative to $k$'s perception of complementarity, then $k$'s contribution to the complementarity of $a$ and $b$ in the presence of $u$, $\Delta^3_{kab}(u)$, is negative even though the goods themselves are still complementary for $k$.

The other interpretation of the third-difference terms in (4.4) is based on the observation that

$$\Delta^3_{kab}(S) \equiv \Delta^2_{ab}(\Delta_k(S)).$$

Thus, $\Delta^3_{kab}(S)$ is positive if and only if the merging players $a$ and $b$ are complementary in their effect on the marginal contribution of player $k$ to coalition $S$. This matters because after merging, $a$ and $b$ contribute or withdraw their joint resources together while before they acted separately, which is what the second difference $\Delta^2_{ab}(S)$ formalizes (as the last line of (4.1) makes clear). If $\Delta^2_{ab}(\Delta_k(S))$ is positive, then the merger of $a$ and $b$ improves the bargaining position of player $k$, which is reflected in an increase in her Shapley value $Sh^k$; if it is negative, $k$'s payoff is reduced by the merger.

The profitability of the merger for the merging players themselves follows from the fact that the worth of the grand coalition $v(N)$ is independent of the

[11]The assumption that player $k$ is essential and $\Delta^2_{ab}(\emptyset) = 0$ is not critical.

[12]Throughout, we write $v(abk)$ for $v(\{a, b, k\})$.

[13]For example, Horn and Wolinsky (1988b) find that a firm benefits when complementary workers form a single union.

allocation of control rights over resources. Hence, the merger results in a zero-sum redistribution of the constant total net surplus, and the benefit to the merging players $a$ and $b$ is equal to the loss of the outsiders in $N \setminus \{a, b\}$:

$$M^{ab}(ab) = -\sum_{k \in N \setminus \{a,b\}} M^k(ab). \tag{4.5}$$

The merger benefit for each firm has a particularly simple form when there are only four players $N = \{a, b, k, j\}$. When $a$ and $b$ merge, the effect on the payoffs of outsiders $k$ and $j$ is, respectively,

$$M^k(ab) = \frac{1}{12}\Big[\Delta^2_{ab}(k) - \Delta^2_{ab}(\emptyset) + \Delta^2_{ab}(kj) - \Delta^2_{ab}(j)\Big] \tag{4.6}$$

and

$$M^j(ab) = \frac{1}{12}\Big[\Delta^2_{ab}(j) - \Delta^2_{ab}(\emptyset) + \Delta^2_{ab}(kj) - \Delta^2_{ab}(k)\Big] \tag{4.7}$$

so that

$$M^{ab}(ab) = \frac{1}{6}\Big[\Delta^2_{ab}(\emptyset) - \Delta^2_{ab}(kj)\Big]. \tag{4.8}$$

For further reference, let

$$R(ab) = \sum_{k \in N \setminus \{a,b\}} 1\left[M^k(ab) < 0\right] M^k(ab)$$

denote the (gross) amount of surplus redistributed by a profitable merger of players $a$ and $b$.

## 4.3 Heterogeneous outsiders

When outsiders are heterogeneous, a merger redistributes total industry surplus among the outsiders as well as between insiders and outsiders. Moreover, the positive effect on one or more outsiders can be arbitrarily large relative to the benefit to the merging players themselves. In contrast with mergers in a Cournot oligopoly à la Salant et al. (1983), this is based on bargaining considerations alone, without any accompanying change in allocation.[14] Moreover, the outsiders need not be competitors to the merging firms; they can be any of the players in a multi-tiered industry connected by active trading links.

[14]Stigler (1950) already notes that outsiders are the main beneficiaries of a merger if they can free-ride on the merging firms' reduction of output. Kamien and Zang (1990) show that this can lead to the failure of the players to form a monopoly and maximize joint profits in a Cournot oligopoly with *endogenous* mergers.

In the literature, differences between outsiders have not received much attention. Outsiders are usually taken as essentially, if not literally, homogeneous. Indeed, in some applications there is only one outsider, so the question of heterogeneity among the outside players does not arise.[15] In the otherwise very general analysis in Segal (2000), the author employs a robustness condition which requires outsiders to be homogeneous with respect to the sign of their effect on the complementarity of the merging firms. This is essentially the same restriction as in Inderst and Wey (2000) where, in a model of horizontal mergers in a two-tiered industry, two suppliers can differ in their cost structures so long as they both have either increasing or decreasing average costs. Similarly, in the technology-choice analysis in Inderst and Wey (2001) only technologies with decreasing unit costs are considered. In Segal's case, the assumption is made to ensure that the sign of $M^{ab}(ab)$ is invariant to the choice of random-order value. Since the set of admissible random-order values in his model includes those which put all probability weight on any single outsider, all outsiders must be qualitatively identical in their effect on $\Delta^2_{ab}(S)$. More precisely, while some differences between the outsiders are allowed, it is required that $\Delta^3_{kab}(S)$ have the same sign for all outsiders $k \in N \setminus \{a, b\}$ and for all $S \subseteq N \setminus \{a, b, k\}$.

For the analysis of more complex real-life situations, the restriction to essentially homogeneous outsiders can be unsatisfying. In particular, intermediate firms in multi-tiered industries bargain with upstream suppliers which are likely to have rather different characteristics than their downstream customers. Even in simple two-sided markets outsiders to a merger can very easily be fundamentally different. In the context of a horizontal downstream merger, for instance, the merging firms might face some upstream suppliers with increasing costs and others with decreasing costs.[16] This is also true for vertical mergers where the outside players belong to different industry tiers. In each case, the requirement that the effect of outsider $i$ on the complementarity of the merging firms $a$ and $b$, $\Delta^3_{iab}(S)$, has the same sign for all $i$ and for all $S$ is quite restrictive.

In contrast with Segal (2000), we take the inherent bargaining abilities of the players as given and known, so we do not consider issues like the invariance of our results to the choice of specific random-order values. Instead, we rely on

[15]For example, in Horn and Wolinsky (1988b) the merger candidates are two groups of workers facing a single firm (the outsider, in our context). The same is true in the section on unionization of Stole and Zwiebel (1996b).

[16]This case is illustrated in the example in Section 4.4.1.

the Shapley value as our solution concept but generalize the analysis by allowing explicitly for heterogeneity among the players, which we define as follows.

**Definition 2 (Heterogeneity)** *Fix two merger candidates $a$ and $b$ in $N$. The outside players in $N \setminus \{a, b, \}$ are called heterogeneous if and only if there are at least two outsiders $i$ and $j$ such that $M^i(ab) > 0$ and $M^j(ab) < 0$.*

This does not imply that the third-difference terms $\Delta^3_{kab}(S)$ which determine $M^k(ab)$ must be of constant sign for each outside player $k \in N \setminus \{a, b\}$ and for all $S \subseteq N \setminus \{a, b, k\}$; our notion of heterogeneity only means that at least one player's third-difference terms are positive, and another player's negative, *on average* over the orderings affected by the merger of $a$ and $b$. In many applications, all the third-difference terms of a given player are either positive or negative. For instance, in the supplier-retailer example in Subsection 4.4.1, they are negative for the increasing-cost supplier and positive for the one producing at decreasing cost. Definition 2 allows for more general cases where the sign of the third-difference terms may depend on the set of preceding players $S$.

Recalling from (4.5) that the profitability of the merger is the negative sum of the merger benefits of all outsiders, it is almost tautological that heterogeneity as defined above is a necessary and sufficient condition for some outsiders to benefit from a profitable merger of players $a$ and $b$. When this is the case, the merger entails a redistribution of surplus not only from outsiders to the merging parties, but also among all outsiders.

Not only can some heterogeneous outsiders benefit from a profitable merger; they can be the main beneficiaries, receiving a larger share of the redistributed surplus than the merging players themselves. This is because the profitability of the merger of $a$ and $b$ to the merging players themselves, $M^{ab}(ab) = -\sum_{k \in N \setminus \{a,b\}} M^k(ab)$ can be close to zero even when the merger benefit $M^k(ab)$ is large and positive for some outsider $k$.

To present this argument in a more precise form, we assume that the characteristic function is continuous in a parameter which affects the underlying individual profit functions of the players in $N$.

**Assumption 7** *The characteristic function $v(S)$ is continuous in some parameter $\lambda \in (\underline{\lambda}, \overline{\lambda}) \subset \mathbb{R}$. There exists a $\lambda_1$ such that $M^{ab}(ab) > 0$, and a $\lambda_0$ such that $M^{ab}(ab) = 0$, where $\lambda_0, \lambda_1 \in (\underline{\lambda}, \overline{\lambda})$.*[17]

[17] For notational simplicity, the parameter $\lambda$ is dropped from all functions building on $v(\cdot)$.

This assumption is sufficient but not necessary for what follows. In particular, the continuity of $v(S)$ with respect to $\lambda$ is analytically convenient but not essential for the effects we describe.

**Proposition 10** *Under Assumption 7, if the outsiders in $N \setminus \{a, b\}$ are heterogeneous given $\lambda_0$, then there is a non-empty range of values for $\lambda$ such that most of the surplus $R(ab)$ redistributed by a profitable merger is received by players in $N \setminus \{a, b\}$. As $\lambda \to \lambda_0$, the share of $R(ab)$ received by one or more outsiders goes to one.*

**Proof.** If, given $\lambda_0$, outsiders are heterogeneous, then there is some outsider $i$ with $M^i(ab) > 0$ when $M^{ab}(ab)$ in (4.5) is zero and $a$ and $b$ are just indifferent about merging. Consequently, all the redistributed surplus goes to outsiders (including, but not necessarily limited to, player $i$). By continuity, there is a range for $\lambda$ close to $\lambda_0$ such that $R(ab) > 2M^{ab}(ab)$. ■

While the condition in Proposition 10 is quite general, to verify its validity one has to compute the outsiders' merger benefits. A simpler sufficient condition is satisfied in many applications where some players typically increase the complementarity of the merging parties more than others do.[18] The following Lemma asserts that the merger benefits of two outsiders $i$ and $j$ differ overall if one of them has a (weakly) larger impact on the complementarity of the merging firms in the presence *of all coalitions containing neither $i$ nor $j$*. Essentially, it is shown that the stronger impact of either player extends to larger preceding coalitions containing the other player, and therefore to all coalitions relevant for the determination of the merger benefits $M^i(ab)$ and $M^j(ab)$ of $i$ and $j$. Note that this does not imply that the third-difference terms of the two players must be of different signs given any specific set of preceding players.

**Lemma 20** *If*

$$\Delta^3_{iab}(S) \geq \Delta^3_{jab}(S) \qquad \forall S \subseteq N \setminus \{a, b, i, j\}, \tag{4.9}$$

*with strict inequality for at least one set $S$, then*

$$M^i(ab) > M^j(ab).$$

[18]In the example in Section 4.4.1, the decreasing-cost supplier always increases the complementarity of the two retailers while the increasing-cost supplier reduces it.

If this holds when the merger candidates are just indifferent about merging, then the outsiders $k \in N \setminus \{a, b\}$ must necessarily be heterogeneous in the sense that some merger benefits $M^k(ab)$ are positive, and some are negative.

**Lemma 21** *If (4.9) holds for $\lambda_0$ (with strict inequality for at least one set $S$), then the outsiders to the merger of $a$ and $b$ are heterogeneous as defined above in the vicinity of $\lambda_0$.*

**Proof.** By definition, $M^{ab}(ab) = -\sum_{k \in N \setminus \{a,b\}} M^k(ab) = 0$ when $\lambda = \lambda_0$. Thus, if there are two outsiders $i$ and $j$ with $M^i(ab) > M^j(ab)$, then there must be some outsider $k$ (not necessarily identical with $i$) with $M^k(ab) > 0$ and another outsider $h$ (not necessarily $j$) with $M^h(ab) < 0$. ■

Consequently, when the condition in Lemma 20 holds around $\lambda_0$, then there is a non-empty range of values of $\lambda$ such that the main benefit of the merger falls on some outsiders.

**Corollary 2** *If (4.9) holds for $\lambda_0$, Proposition 10 applies.*

Note, however, the condition that $M^{ab}(ab)$ must be zero for some $\lambda_0$. Without this condition, Segal's requirement that $\Delta^3_{iab}(S) < 0$ for all $i$ and $S$ would not be in conflict with (4.9). However, when all third-difference terms are negative for all outsiders, then there is no $\lambda_0$ such that $M^{ab}(ab)$ is zero. While the continuity of the characteristic function with respect to $\lambda$ is not necessary, the existence of $\lambda_0$ is necessary for (4.9) to be a sufficient condition for Proposition 10.

## 4.4 Shocks

In this section we are more explicit about the kind of exogenous parameter change considered in the previous section. We continue to take two merger candidates $a$ and $b$ from $N$ as exogenously given and concentrate on a shock specific to one of the outsiders in $N \setminus \{a, b\}$.

**Definition 3 ($k$-specific shock)** *A shock is specific to player $k$ if it affects $v(S)$ only if $k \in S$. The shock is called positive if $v(S)$ is weakly increased for all $S \subseteq N$. It is called negative if $v(S)$ is weakly reduced for all $S \subseteq N$.*

As an example, one might think of a decrease in the marginal cost of a supplier due to technological progress, or a decline in demand in a market where one of the players is a local monopolist. We take this shock to be permanent while mergers are not, so that our merger candidates can change their merger or de-merger decision after observing the shock to outsider $k$. This makes our approach complementary to the technology-choice analysis in Inderst and Wey (2001) where the players can switch their production technologies to improve their bargaining position (possibly at the expense of overall efficiency) under a constant market structure. In our model, the shock observed before the merger decision might also result from player $k$'s choice of technology or R&D investment; what matters is only that the shock is taken as given when the merger decision is made.

It turns out that in this setting the final payoff effect of a shock can be very different from the initial effect when the players are heterogeneous and the shock triggers a merger or de-merger of $a$ and $b$. This has some important implications for technology choice and R&D incentives, as well as for empirical work.

A $k$-specific shock directly affects player $k$'s payoff $Sh_k$, since his Shapley value is a sum of first-difference terms $\Delta_k(S) \equiv v(S \cup k) - v(S)$, $k \notin S$, where by definition only the first term on the RHS is affected by the shock. In addition, the shock has an impact on the worth of the grand coalition $v(N)$. Thus, a negative shock to player $k$ has a negative direct effect on both $k$'s payoff and the total surplus available to all players.

Moreover, a $k$-specific shock has an impact on the merger benefits $M^i(ab)$ received by all outsiders $i \in N \setminus \{a, b\}$ when $a$ and $b$ merge, and hence on the profitability of the merger itself. This is because $M^i(ab)$ is the sum of third-difference terms $\Delta^3_{iab}(B'(i, \pi))$ measuring the impact of player $i$ on the complementarity of $a$ and $b$ in the presence of the preceding players $B'(i, \pi)$. For $i = k$, the effect of a $k$-specific shock is obvious; for players other than $k$, the third-difference terms in $M^i(ab)$ are affected because $k$ appears in some of the preceding coalitions $B'(i, \pi)$ over which $M^i(ab)$ is evaluated.[19]

This is particularly relevant when the outside players are heterogeneous and the merger decision involves a trade-off between extracting a larger share of the total surplus from some outsiders and losing more of it to others. If a $k$-specific shock affects the outsiders' merger benefits sufficiently to change the sign of the merger profitability $M^{ab}(ab)$, then it leads to an additional, indirect effect on the payoffs of all players, including $k$ itself, which may well can go in the opposite direction of the original shock. Indeed, the merger effect can overcompensate the

[19] In the four-firm example in (4.6) and (4.7), a shock to $k$ affects both $\Delta^2_{ab}(k)$ and $\Delta^2_{ab}(kj)$ and, therefore, $M^j(ab)$.

direct shock, leading to the opposite overall payoff effect that one would expect from observing the $k$-specific shock alone.

**Proposition 11 (Merger compensates negative shock)** *Consider a negative $k$-specific shock with a continuous effect measured by some parameter $\lambda \in \mathbb{R}$. Suppose that before the shock sign$(M^{ab}(ab)) = -$sign$(M^k(ab))$ and there is a $\lambda_0$ such that the sign of $M^{ab}(ab)$ changes into that of $M^k(ab)$, which is bounded away from zero. Then the payoff of player $k$ is decreasing in the size of the shock except for a discontinuity (an upward jump) at $\lambda_0$. Moreover, if $\lambda$ was sufficiently close to $\lambda_0$ before the shock, then player $k$ is better off overall.*

The merger decision of $a$ and $b$ is a trade-off in bargaining positions vis-à-vis different trading partners. Those who enhance the complementarity of $a$ and $b$ give them an incentive to remain separate; those who reduce it make a merger more profitable. If $M^{ab}(ab)$ and $M^k(ab)$ are of opposite signs prior to a shock but have the same sign afterwards, then player $k$ benefits from a change in the merger decision. A negative $k$-specific shock has a negative direct effect on $k$'s payoff, but if it reverses the sign of $M^{ab}(ab)$, some or all of the initial effect is compensated.

The discontinuity in $k$'s total payoff is due to the fact that Shapley values and merger benefits are continuous in $\lambda$ but the merger decision is discrete and $M^k(ab)$ is bounded away from zero by assumption. As argued in the previous section, the heterogeneity of the outside players as defined in Definition 2 is a necessary condition for a merger to compensate part of the negative shock to $k$, or even to overcompensate it. The flip side of this observation is that there must be some players from whom the surplus redistributed to $k$ is taken. Since the negative shock to $k$ reduces the worth of the grand coalition $v(N)$, those players receive an even smaller share of a reduced total surplus while player $k$, who was originally hit by the negative shock, is left better off than before. In Section 4.4.1 we illustrate this point in a simple example of a two-tiered producer/retailer industry where a supplier operating at decreasing costs suffers a negative cost shock whose final impact is borne primarily by another (and only indirectly connected) player.

**Corollary 3 (Merger compensates positive shock)** *Suppose the $k$-specific shock in Proposition 11 is positive instead of negative and sign$(M^{ab}(ab)) =$ sign$(M^k(ab))$ before the shock. If there is a $\lambda_0$ such that the sign of $M^{ab}(ab)$ changes into the opposite of sign$(M^k(ab))$, then the total payoff of player $k$ is increasing in the size of the shock except for a downward jump at $\lambda_0$. If $\lambda$ was sufficiently close to $\lambda_0$ before the shock, then player $k$ is worse off overall.*

It is straightforward to extend this to a setting where the shock is created by one of the players rather than imposed exogenously. In particular, a $k$-specific shock might be the result of player $k$'s technology choice or R&D effort prior to the merger decision of $a$ and $b$.[20] The compensating effect described in Proposition 11 and its corollary implies a strategic incentive for player $k$ to choose an inferior technology if this induces a merger of $a$ and $b$ improving his bargaining position sufficiently to make up for the loss of productive efficiency. Likewise, player $k$ will strategically underinvest in R&D if the efficient research effort would result in an adverse merger of $a$ and $b$, preventing $k$ from fully appropriating the gains from his efforts.

For empirical studies, the compensating effect of mergers implies that the correlation between firm-specific shocks and profits can be distorted, or even reversed, when the trading partners of the affected firm merge or de-merge. In contrast with mergers in Cournot models à la Salant et al. (1983), such mergers need not take place on the same industry tier. As a real-world example, consider the widely-reported demand stagnation in the automobile industry during the early 1990s which was accompanied – accidentally or not – by the formation of "module" suppliers bundling complementary inputs for the final downstream car manufacturers. By casual observation, the negative demand shock has failed to translate into lower profits for manufacturers. From a bargaining perspective, this is not surprising since the merger of complementary suppliers should benefit their downstream customers. This begs the question, however, why those complementary suppliers would merge in the first place. Propositions 10 and 11 provide an explanation: If the merging suppliers face some other, fundamentally different, trading partners in addition to final manufacturers (e. g. their own upstream suppliers), then the loss of bargaining power vis-à-vis the car makers can be outweighed by an improved bargaining position in their other negotiations. The main impact of the negative shock to car makers may then fall on firms which are only indirectly linked to them.

### 4.4.1 Example: Horizontal merger in a retailing model

Consider the horizontal-merger decision of two retailers $a$ and $b$ facing two producers $c$ and $d$ in a four-firm industry à la Inderst and Wey (2001) illustrated

[20] This assumes that the choice of technology cannot be reversed after the merger. See Inderst and Wey (2001) for the opposite case.

in Figure 4.1. Let $x_{ac}$ denote the quantity supplied to retailer $a$ by producer $c$,

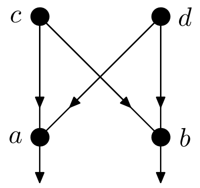


**Figure 4.1:** Two retailers $a$ and $b$ selling the output of both upstream producers $c$ and $d$ in separate local markets.

and analogously for the other three trading links. The cost function of producer $i = c, d$ is given by

$$\kappa_i(x_{ai} + x_{bi}) + \delta_i(x_{ai} + x_{bi})^2, \qquad \kappa_i \geq 0.$$

To introduce the kind of fundamental heterogeneity defined above, we assume that $\delta_c > 0$ and $\delta_d \in (-\frac{1}{2}, 0)$, implying increasing unit costs for producer $c$ and decreasing costs for producer $d$. Each retailer $i = a, b$ operates as a local monopolist, achieving a gross profit (before payments to the upstream suppliers) of

$$\alpha(x_{ic} + x_{id}) - (x_{ic}^2 + x_{id}^2), \qquad \alpha > \max\{\kappa_c, \kappa_d\}$$

from selling the two goods.[21] At least one retailer and one producer must be present for a coalition of firms to achieve a positive profit; this implies that $\Delta_{ab}^2(\emptyset)$ is zero. Hence, if $a$ and $b$ merge, then the merger benefits to the outsiders $c$ and $d$ are[22]

$$M^c(ab) = \frac{1}{12}\left(\Delta_{ab}^2(c) - \Delta_{ab}^2(d) + \Delta_{ab}^2(cd)\right) \tag{4.10}$$

and

$$M^d(ab) = \frac{1}{12}\left(\Delta_{ab}^2(d) - \Delta_{ab}^2(c) + \Delta_{ab}^2(cd)\right), \tag{4.11}$$

while the merging firms improve their joint payoff by

$$M^{ab}(ab) = -\frac{1}{6}\Delta_{ab}^2(cd).$$

[21]The gross-profit function implies that the goods are neither (strict) substitutes nor complements, and that the retailers have no operating cost. Neither assumption is critical. The local-monopoly assumption conveniently eliminates coordination problems for $a$ and $b$ vis-à-vis final consumers (but not against $c$ and $d$).

[22]Cf. (4.6)–(4.8)

Maximizing joint profits over the quantities traded among each coalition of firms gives

$$\begin{aligned}\Delta^2_{ab}(c) &= v(abc) - v(ac) - v(bc)\\ &= -\frac{\delta_c(\alpha-\kappa_c)^2}{2(1+\delta_c)(1+2\delta_c)},\end{aligned}$$

$$\begin{aligned}\Delta^2_{ab}(d) &= v(abd) - v(ad) - v(bd)\\ &= -\frac{\delta_d(\alpha-\kappa_d)^2}{2(1+\delta_d)(1+2\delta_d)},\end{aligned}$$

and

$$\begin{aligned}\Delta^2_{ab}(cd) &= v(abcd) - v(acd) - v(bcd)\\ &= -\frac{\delta_c(\alpha-\kappa_c)^2}{2(1+\delta_c)(1+2\delta_c)} - \frac{\delta_d(\alpha-\kappa_d)^2}{2(1+\delta_d)(1+2\delta_d)}\\ &= \Delta^2_{ab}(c) + \Delta^2_{ab}(d).\end{aligned}$$

(The equality of $\Delta^2_{ab}(cd) = \Delta^2_{ab}(c) + \Delta^2_{ab}(d)$ stems from the fact that the downstream goods are neither complements nor substitutes.)

Since $\delta_c > 0$ and $\delta_d < 0$ by assumption, we have $\Delta^2_{ab}(c) < 0$ and $\Delta^2_{ab}(d) > 0$. By (4.10), (4.11) and (4.12), the merger benefit is negative for the increasing-cost outsider $c$ and positive for outsider $d$: $M^c(ab) = 2\Delta^2_{ab}(c) < 0$ and $M^d(ab) = 2\Delta^2_{ab}(d) > 0$. Thus, the upstream firms are heterogeneous according to Definition 2, and they satisfy condition (4.9) for all admissible values of $\alpha$, $\kappa_c$ and $\kappa_d$.[23] As a consequence of Corollary 2, when the merger incentive for $a$ and $b$ is close to zero, one of the outsiders receives most of the redistributed surplus. Indeed, one can easily find parameter values such that $\Delta^2_{ab}(d) = -\Delta^2_{ab}(c) > 0$, in which case $a$ and $b$ are indifferent about merging, and the merger effectively redistributes surplus from the increasing-cost producer $c$ to the supplier $d$ operating a decreasing unit cost.

Now suppose that initially the merger candidates $a$ and $b$ prefer to stay apart, i. e. $M^{ab}(ab) < 0$ because $\Delta^2_{ab}(cd) = \Delta^2_{ab}(c) + \Delta^2_{ab}(d) > 0$. Suppose further that producer $d$ is hit by a negative cost shock increasing $\kappa_d$ by $\lambda$, which, other things equal, has a continuous, negative impact on $d$'s Shapley value $Sh^d$. At the same time, the shock leaves $\Delta^2_{ab}(c) < 0$ unchanged but decreases $\Delta^2_{ab}(d)$ arbitrarily close

[23] $\Delta^2_{ab}(d) > \Delta^2_{ab}(c)$ implies $\Delta^3_{dab}(\emptyset) \equiv \Delta^2_{ab}(d) - \Delta^2_{ab}(\emptyset) > \Delta^3_{cab}(\emptyset) \equiv \Delta^2_{ab}(c) - \Delta^2_{ab}(\emptyset)$ for all $\kappa_c < \alpha$ and $\kappa_d < \alpha$.

to zero.[24] Thus, player $d$ increases the complementarity between $a$ and $b$ by less than before the shock. This affects the trade-off for the merger candidates $a$ and $b$ between extracting more rent from $c$ by merging and from $d$ by staying separate, and if $\lambda$ is large enough, then $a$ and $b$ reverse their initial decision and merge. As a function of $\lambda$, therefore, player $d$'s total payoff is non-monotonic. Precisely, it is decreasing except for an upward jump at the point where $\Delta^2_{ab}(d) = -\Delta^2_{ab}(c)$, at which the merger becomes profitable. If $\Delta^2_{ab}(d)$ was close to $-\Delta^2_{ab}(c)$ before the shock and the shock is not too large, then by the continuity of $v$, the negative effect on $Sh^d$ is less than the merger benefit $M^d(ab)$, and player $d$ benefits overall from the negative shock, while player $c$ suffers (since $M^c(ab) < 0$). The negative effect of a shock to supplier $d$ is borne primarily by the increasing-cost supplier $c$ which is connected to $d$ only through their common downstream customers.

Instead of being subject to an exogenous shock, player $d$ might have to make a binding technology choice before $a$ and $b$ decide about merging. The range of negative-shock sizes leaving $d$ better off overall then translates into a range of technology parameters such that $d$'s final payoff is maximized by choosing an inferior technology lest $\Delta^2_{ab}(d)$ be so large as to induce a merger of $a$ and $b$. By the same line of reasoning, a *positive* shock specific to player $d$ can be compensated by an *adverse* merger decision by $a$ and $b$, leaving $d$ worse off than before the shock. This implies a disincentive to engage in cost-reducing R&D for a player with a positive merger-benefit term $M^k(ab)$, i. e. for a trading partner which enhances the complementarity of $a$ and $b$. For instance, the incentive for a decreasing-cost supplier to reduce its marginal cost is inefficiently low if one of its customers responds by de-merging into two independent units.

## 4.5 Interaction between merger decisions

So far we have only considered a single merger of two predetermined players. In this section, we analyze the effect of one merger on the profitability of another, where both pairs of merger candidates are exogenously fixed. While this is still much less ambitious than allowing for the endogenous determination of merging partners, we believe it to be a significant building block for a more complete theory of bargaining and mergers, and for a better understanding of "merger waves".

[24]The derivative of $\Delta^2_{ab}(d)$ w. r. t. $\kappa_d$ is negative for $\kappa_d \in [0, \alpha)$, and $\Delta^2_{ab}(d) \to 0$ as $\kappa_d \to \alpha$.

As a stylized empirical fact, it is often observed that mergers occur in waves. There are several possible explanations for this, from economy-wide shocks increasing the profitability of mergers (like changes in the tax code or in the antitrust regime) to herding effects and management fads. Nilssen and Sørgard (1998) show how mergers can be strategic complements in a Cournot model à la Salant et al. (1983). While this suggests that market power considerations might lead to an interaction of merger incentives for disjoint sets of players, this is not disentangled from the monopolization effect and the cost synergies assumed by the authors. On the other hand, Inderst and Wey (2001) find that merger incentives are independent in a model based purely on bargaining effects.

The results in this section show that even in a pure bargaining model, the decisions of two disjoint pairs of merger candidates in an $n$-firm industry generally interact. Only in the special case of $n = 4$, when there are no "complete" outsiders who do not participate in either merger, are pairwise mergers independent of each other. Generally, the merger of two firms $a$ and $b$ affects the impact $\Delta^3_{kcd}(S)$ of each "complete" outsider $k \in N \setminus \{a, b, c, d\}$ on the complementarity of players $c$ and $d$ and, therefore, the amount of surplus $M^k(cd)$ received by those outsiders when $c$ and $d$ merge. This is because some of the preceding coalitions $S$ contain $a$ and $b$, who after their merger appear either jointly or not all where before they appeared separately. As will be shown, this leads to a symmetric interaction between the merger incentives for $a$ and $b$ on the one hand, and $c$ and $d$ on the other. The interaction can be positive, in which case mergers are strategic complements; if it is negative, they are strategic substitutes.

Let $a$ and $b$ denote two players pondering a merger, and let $c$ and $d$ denote a disjoint second pair of potential merger candidates, so that $\{a, b\} \cap \{c, d\} = \emptyset$ and $\{a, b\} \cup \{c, d\} \subseteq N$. Only the $a$–$b$ merger and the $c$–$d$ merger are allowed; in particular, the firms cannot form a grand merger containing all four of them. We first look for the effect of a merger of $a$ and $b$ on the private profitability $M^{cd}(cd)$ of the $c$–$d$ merger. This is the same as the negative sum of the effects of the $a$–$b$ merger on the merger benefits $M^k(cd)$ received by all outsiders $k \in N \setminus \{c, d\}$ when $c$ merges with $d$. Among the outsiders to the second merger are the players who have formed the first; of these, $a$ is again, and without loss of generality, treated as the proxy player while $b$ is taken to be a dummy. The $a$–$b$ pair needs to be considered separately from those players $k \notin \{a, b, c, d\}$ who participate in neither merger, because after their own merger, $a$ and $b$ contribute a different set of marginal resources to the same preceding coalitions, while the "complete" outsiders $k \in N \setminus \{a, b, c, d\}$ contribute the same resources as before to different preceding coalitions.

The two types of outsiders are considered in turn before the total effect of the $a$–$b$ merger on $M^{cd}(cd)$ is stated in Proposition 12.

### 4.5.1 Complete outsiders

We begin with the "complete" outsiders $k \notin \{a, b, c, d\}$, restricting our attention to those orderings where $c \prec_\pi k \prec_\pi d$, i. e. $k$ follows $c$ and precedes $d$, which are the only relevant ones for

$$M^k(cd) = \frac{1}{n!} \sum_{\pi \in \Pi} 1[c \prec_\pi k \prec_\pi d] \Delta_{kcd}(B'(k, \pi)), \tag{4.12}$$

where

$$B'(k, \pi) \equiv B(k, \pi) \setminus \{c, d\},$$

like before, stands for the set of outsiders to the $c$–$d$ merger who precede player $k$ in ordering $\pi$.

If both $a$ and $b$ precede $k$ in ordering $\pi$, then the $a$–$b$ merger has no relevant effect on the $k$-preceding coalition $B'(k, \pi)$.[25] The same is true when both $a$ and $b$ appear after $k$. This leaves those orderings where $k$ comes between $\{a, b\}$ and $\{c, d\}$.

If $\{a, c\} \prec_\pi k \prec_\pi \{b, d\}$, then after the $a$–$b$ merger, the relevant preceding coalition is $B'(k, \pi) \cup b$ instead of just $B'(k, \pi)$. Conversely, when $\{b, c\} \prec_\pi k \prec_\pi \{a, d\}$, the post-merger preceding coalition is $B'(k, \pi) \setminus b$ rather than $B'(k, \pi)$. These two cases can be pairwise combined and aggregated quite concisely with the help of some additional notation.

Define the set of preceding players outside of both merger pairs as

$$\begin{aligned} B''(k, \pi) &\equiv \{i \in N \setminus \{a, b, c, d\} : i \prec_\pi k\} \\ &\equiv B'(k, \pi) \setminus \{a, b\}. \end{aligned} \tag{4.13}$$

For each ordering $\pi \in \Pi$ with $a \prec_\pi k \prec_\pi b$ there is exactly one corresponding ordering $\pi^* \in \Pi$ which differs only in the transposition of $a$ and $b$. In particular, the set of "complete" outsiders preceding $k$ is the same in both orderings,

[25]Recall that only the identity of the players in $B'(k, \pi)$ matters for $\Delta^3_{kab}(B'(k, \pi)$; their ordering *within* $B'(k, \pi)$ is irrelevant. Moreover, adding the dummy player to $B'(k, \pi)$ has no effect.

$B''(k,\pi) = B''(k,\pi^*)$. Thus, the effects in the two relevant types of orderings can written as

$$\frac{1}{n!}\sum_{\pi\in\Pi} 1[\{a,c\} \prec_\pi k \prec_\pi \{b,d\}]\Big[\Delta^3_{kcd}(B''(k,\pi)\cup\{a,b\}) + \Delta^3_{kcd}(B''(k,\pi)) - \Delta^3_{kcd}(B''(k,\pi)\cup a) - \Delta^3_{kcd}(B''(k,\pi)\cup b)\Big]. \tag{4.14}$$

This is the average complementarity of $a$ and $b$ as members of the coalition $B'(k,\pi)$ preceding $k$, for $\Delta^3_{kcd}(B'(k,\pi))$. Intuitively, the complementarity of $a$ and $b$ matters because after their merger, they appear either in combination or not at all, where before they would contribute their resources separately. This is emphasized in the following representation of (4.14) using the second-difference operator $\Delta^2_{ab}$.

**Lemma 22** *The effect of the a–b merger on the c–d merger benefit for each "complete" outsider $k \in N \setminus \{a,b,c,d\}$ in (4.14) is equivalent to*

$$\frac{1}{n!}\sum_{\pi\in\Pi} 1[\{a,c\} \prec_\pi k \prec_\pi \{b,d\}]\Delta^2_{ab}\Big(\Delta^2_{cd}\left[B''(k,\pi)\cup k\right] - \Delta^2_{cd}\left[B''(k,\pi)\right]\Big) \tag{4.15}$$

This follows directly from the definition of the difference operators $\Delta^2_{ab}$ and $\Delta^3_{kcd}$.

The next step is to aggregate the effect in (4.15) across all "complete" outsiders $k \in N \setminus \{a,b,c,d\}$. To do this, we can exploit the fact that (4.15), and hence the sum over all outsiders, consists of terms of the form $\Delta^2_{ab}(\Delta^2_{cd}(S))$, where $S \subseteq N \setminus \{a,b,c,d\}$. Therefore, the total effect on all complete outsiders can be computed by adding over all subsets $S \subseteq N \setminus \{a,b,c,d\}$, which leads to a particularly simple expression.

**Lemma 23** *Across all "complete" outsiders $k \in N \setminus \{a,b,c,d\}$, the effect of the a–b merger on the benefits $M^k(cd)$ received from the c–d merger is*

$$-\frac{2}{n!}\sum_{S\subseteq N\setminus\{a,b,c,d\}} \alpha_{|S|}\Delta^2_{ab}\left[\Delta^2_{cd}(S)\right], \tag{4.16}$$

*where the coefficient*

$$\alpha_{|S|} = (n-2|S|-4)\,(|S|+1)!\,(n-|S|-3)! \tag{4.17}$$

*depends only on $n$ and the cardinality of $S$.*

We postpone the discussion of $\alpha_{|S|}$, which has some interesting characteristics, and of the nested second differences, until Proposition 12 has been derived.

### 4.5.2 Merging outsiders

The merging players $a$ and $b$ need to be considered separately as outsiders to the merger of $c$ and $d$ because they change their marginal contribution to any given preceding coalition. Consider first the dummy player $b$. If $a$ and $b$ stay apart, then the effect of the $c$–$d$ merger on the payoff of $b$ is

$$M^b(cd) = \frac{1}{n!} \sum_{\hat{\pi} \in \Pi} 1[c \prec_{\hat{\pi}} b \prec_{\hat{\pi}} d] \Delta^3_{bcd}(B'(b, \hat{\pi})). \tag{4.18}$$

In contrast, when $a$ and $b$ have merged and $b$ has become a dummy player, the effect of the $c$–$d$ merger on $b$ is zero by definition. Thus, by participating in the first merger with $a$, player $b$ loses $\Delta^3_{bcd}(B'(b, \hat{\pi}))$ in each ordering $\hat{\pi}$ with $c \prec_{\hat{\pi}} b \prec_{\hat{\pi}} d$. (The loss may well be negative, i. e. a gain.)

The proxy player $a$, on the other hand, receives a $c$–$d$ merger benefit (if positive; damage if negative) of

$$M^a(cd) = \frac{1}{n!} \sum_{\pi \in \Pi} 1[c \prec_{\pi} a \prec_{\pi} d] \Delta^3_{acd}(B'(a, \pi)) \tag{4.19}$$

if $a$ and $b$ are separate players, and

$$M^{(ab)}(cd) = \frac{1}{n!} \sum_{\pi \in \Pi} 1[c \prec_{\pi} a \prec_{\pi} d] \Delta^3_{(ab)cd}(B'(a, \pi)) \tag{4.20}$$

when they are merged, where the $(ab)$ notation indicates that $a$ now controls the joint resources of $a$ and $b$.

Our objective for the moment is to determine how the combined $c$–$d$ merger benefits for $a$ and $b$ change when $a$ and $b$ merge themselves. Formally, we look for

$$M^{(ab)}(cd) - M^a(cd) - M^b(cd). \tag{4.21}$$

Since $M^{(ab)}(cd)$, $M^a(cd)$, and $M^b(cd)$ contain different third-difference terms evaluated over various preceding coalitions in different subsets of $\Pi$, solving for the aggregate effect in (4.21) involves a number of steps. First, the fundamental symmetry of the permutations in $\Pi$ allows us to sum the terms in $M^b(cd)$ over the equivalent orderings with $c \prec_{\pi} a \prec_{\pi} b$ obtained by the transposition of $a$ and $b$ in each ordering $\hat{\pi}$ in (4.18). Second, the third-difference terms $\Delta^3_{icd}(B'(i, \pi))$ can be expressed in terms of second differences $\Delta^2_{cd}(\cdot)$ and sets of complete outsiders $B''(a, \pi)$ preceding $a$, so that each of the merger benefits in (4.21) can be written as a sum of comparable terms. The following lemma describes the result of these operations. The proof in Section 4.7.2 in the Appendix contains the details.

**Lemma 24**

$$M^{(ab)}(cd) - M^a(cd) - M^b(cd) = \\ \frac{1}{n!}\sum_{\pi\in\Pi}\Big\{1\Big[c \prec_\pi a \prec_\pi \{b,d\}\Big]\Delta^2_{ab}\left(\Delta^2_{cd}[B''(a,\pi)]\right) \\ - 1\Big[\{b,c\} \prec_\pi a \prec_\pi d\Big]\Delta^2_{ab}\left(\Delta^2_{cd}[B''(a,\pi)]\right)\Big\}. \quad (4.22)$$

As before, this expression can be rewritten as the sum of terms of the form $\Delta^2_{ab}(\Delta^2_{cd}(S))$ over all coalitions $S \subseteq N \setminus \{a,b,c,d\}$ of preceding outsiders.

**Lemma 25** *The effect in (4.22) of the merger of players a and b on their own benefit derived from a merger of players c and d is equivalent to*

$$\frac{1}{n!}\sum_{S\subseteq N\setminus\{a,b,c,d\}} \alpha_{|S|}\Delta^2_{ab}\left[\Delta^2_{cd}(S)\right], \quad (4.23)$$

*with* $\alpha_{|S|}$ *as defined in (4.17).*

### 4.5.3 Total effect of the first merger on the second

The total effect of the merger of $a$ and $b$ on the profitability of the $c$–$d$ merger is the negative sum of (4.16) and (4.23).

**Proposition 12** *The effect of the merger of a and b on the profitability* $M^{cd}(cd)$ *of the c–d merger is given by*

$$\frac{1}{n!}\sum_{S\subseteq N\setminus\{a,b,c,d\}} \alpha_{|S|}\Delta^2_{ab}\left[\Delta^2_{cd}(S)\right], \quad (4.24)$$

*with* $\alpha_{|S|}$ *as defined in (4.17).*

A number of points are worth noting about the coefficients $\alpha_{|S|}$. First, they depend only on $n$ and on the size of each coalition $S \subseteq N \setminus \{a,b,c,d\}$. Second, they are positive for $S = \emptyset$, negative for $S = N \setminus \{a,b,c,d\}$, and non-increasing in $|S|$. Third, they are symmetric in the sense that

$$\alpha_{|S|} = -\alpha_{n-|S|-4},$$

so that for each $S$ of size $|S|$ there is a corresponding $S' \subseteq N \setminus \{a,b,c,d\}$ of size $n - |S| - 4$ whose coefficient is minus that of $S$.

For example, for $n = 5$ and $N = \{a, b, c, d, e\}$, (4.24) becomes

$$\begin{aligned} &\frac{1}{120}\Big(2\Delta^2_{ab}\left[\Delta^2_{cd}(\emptyset)\right] - 2\Delta^2_{ab}\left[\Delta^2_{cd}(e)\right]\Big) \\ =& -\frac{2}{120}\Delta^2_{ab}\left[\Delta^3_{ecd}(\emptyset)\right] \\ =& -\frac{2}{120}\Delta^2_{ab}\left[\Delta^2_{cd}(\Delta_e(\emptyset))\right] \end{aligned} \tag{4.25}$$

Hence, the $a$–$b$ merger increases the profitability of the $c$–$d$ merger if and only if $a$ and $b$ are non-complementary in their effect on $\Delta^3_{ecd}(\emptyset)$. Recall that when $c$ and $d$ merge, some of the redistributed surplus goes to $e$ if $\Delta^3_{ecd}(\emptyset) > 0$, i. e. if $e$ enhances the complementarity of $c$ and $d$. In order to make the $c$–$d$ merger more attractive, the merger of $a$ and $b$—in essence, the decision to act in combination rather than separately—must *reduce* this third-difference term, which is just what (4.25) implies.[26]

When $n > 5$, (4.24) expands to a sum of terms of the form

$$-\Delta^2_{ab}\Big[\Delta^2_{cd}(S') - \Delta^2_{cd}(S)\Big],$$

where $S$ is a strict subset of $S'$. The term in square brackets describes the impact of the players in $S' \setminus S$ on the complementarity of the merging parties $c$ and $d$; this corresponds to the third-difference terms in (4.12) describing the impact of a single outsider on the complementarity of the merging firms. Hence, it is a measure of the amount of surplus redistributed to outsiders when $c$ and $d$ merge. To increase the profitability of the merger of $c$ and $d$, the first merger of $a$ and $b$ must reduce the amount of surplus going to outsiders. It does so if and only if $a$ and $b$ are non-complementary with respect to the term in square brackets, i. e. if and only if $\Delta^2_{ab}[\,\cdot\,]$ is negative.

Among other things, Proposition 12 implies that there is no effect of the first merger on the second when there are no other players besides $\{a, b, c, d\}$. In contrast, the two merger decision are generically interdependent when $n > 4$. Thus, while four-player models might appear to be a natural starting point for the analysis of two disjoint mergers, their results do not necessarily generalize to a setting with $n > 4$.

**Corollary 4 (n = 4)** *Disjoint pairwise mergers are independent if $n = 4$. They are generically interdependent if $n > 4$.*

[26] Cf. the discussion in Section 4.5.4 of the solid curve in Figure 4.3

Another implication of (4.24) is the fundamental symmetry of the interaction between disjoint mergers.

**Corollary 5 (Symmetry of merger interaction)** *The impact of a merger of $a$ and $b$ on the profitability of a merger between $c$ and $d$ is the same as the effect of the $c$–$d$ merger on the profitability of the merger of $a$ and $b$.*

This symmetry follows from the fact that $\Delta^2_{ab}(\Delta^2_{cd}(S)) = \Delta^2_{cd}(\Delta^2_{ab}(S))$, i. e. that the order of taking differences over $v(S)$ is irrelevant. Abusing terminology, one can think of $\Delta^2_{cd}(S)$ as the first derivative of the characteristic function with respect to the (binary) merger decision of $c$ and $d$, evaluated at $S$. By the same token, $\Delta^2_{ab}(\Delta^2_{cd}(S))$ corresponds to the cross-derivative of $v$ with respect to both mergers, so that the symmetry of interaction between them is not surprising.

Proposition 13 follows directly.

**Proposition 13** *Merger decisions are strategic complements if (4.24) is positive. A sufficient condition is that $\Delta^2_{ab}(\Delta^2_{cd}(S))$ is non-increasing in the cardinality of $S$ for all $S \subseteq N \setminus \{a, b, c, d\}$.*

*Conversely, mergers are strategic substitutes if (4.24) is negative, for which it is sufficient that $\Delta^2_{ab}(\Delta^2_{cd}(S))$ is non-decreasing in the cardinality of $S$ for all $S \subseteq N \setminus \{a, b, c, d\}$.*

The intuition behind the sufficient conditions in Proposition 13 is that, if $\Delta^2_{ab}(\Delta^2_{cd}(S))$ is non-increasing in $S$, then $\Delta^2_{ab}(\Delta^3_{kcd}(S)) \equiv \Delta^2_{ab}(\Delta^2_{cd}(S \cup k) - \Delta^2_{cd}(S))$ is non-positive for all outsiders $k \in N \setminus \{a, b, c, d\}$, $k \notin S$. Thus, the third-difference terms which sum to $k$'s merger benefit $M^k(cd)$ in (4.12) are weakly reduced by the merger of $a$ and $b$, and hence the share of surplus received by $c$ and $d$ after their merger increases. Conversely, if $\Delta^2_{ab}(\Delta^2_{cd}(S))$ is non-decreasing in $S$, then the $a$–$b$ merger increases $M^k(cd)$ for all outsiders $k$, reducing the incentive for $c$ and $d$ to merge as well.

The fact that merger decisions can be either strategic substitutes or complements gives rise to a familiar kind of considerations.[27] In particular, the interaction of merger decisions implies a rationale for merger waves based entirely on bargaining considerations. When two pairwise mergers are unprofitable in isolation but profitable if undertaken simultaneously, then the removal of an obstacle to one of them (an easing of sector-specific regulation, say) can trigger a merger of firms which, prima facie, should not have been affected by that obstacle before.

[27]See, e. g., Bulow et al. (1985). Nilssen and Sørgard (1998) apply the "puppy dog/fat cat" framework of Fudenberg and Tirole (1984) to the analysis of sequential mergers in a Cournot oligopoly.

### 4.5.4 Example: Vertical mergers in a three-tiered industry

As an example for the interaction between disjoint pairwise mergers, consider the simple extension of the Inderst–Wey retailing model from section 4.4.1 illustrated in Figure 4.2. As before, there are two local-monopoly retailers selling the goods produced by two suppliers. For consistency with the notation used in section 4.5, let $b$ and $d$ denote the retailers, and let $a$ and $c$ denote their direct suppliers, each of whom produces a single good. In addition, there is now an upstream firm $e$ producing an intermediate input good used by $a$ and $c$. We look for the interaction between the vertical mergers of $a$ and $b$ on the one hand, and $c$ and $d$ on the other.

By Proposition 13 and (4.25), the two merger decisions are strategic complements if the $a$–$b$ merger reduces the marginal contribution of the complete outsider $e$ to the complementarity of players $c$ and $d$, and strategic substitutes otherwise. This is the case when the goods produced by suppliers $a$ and $c$ are sufficiently complementary; when they are not, the mergers are strategic substitutes. The final outcome is a segmentation of the range of the substitutability parameter $\gamma$ into cases where one or both pairs of merger candidates merge depending on the action taken by the other.

In the following we specify our example and solve for the interaction effect described in Proposition 12, before discussing in detail both the strategic effect and merger incentives for a single pair in isolation.

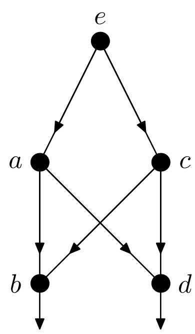


**Figure 4.2:** Two retailers $b$ and $d$, two intermediate-level firms $a$ and $c$, and an upstream producer $e$.

Let the gross profit of each retailer $i = b, d$ be given by

$$\alpha(x_{ia} + x_{ic}) - (x_{ia}^2 + x_{ic}^2 + 2\gamma x_{ia} x_{ic}),$$

where $x_{ia}$ and $x_{ic}$ is the quantity supplied to retailer $i$ by the intermediate-level firms $a$ and $c$, respectively, and where $\gamma \in [-1, 1]$ measures the degree of complementarity between the two goods as perceived by the final customers patronizing each retailer; the goods are complements if $\gamma < 0$ and substitutes if $\gamma > 0$.

Let the intermediate-level firms $a$ and $c$ produce each unit of their differentiated goods using as their only input one unit of the homogeneous output of the upstream firm $e$. Firm $e$, in turn, produces the total quantity $x_{ae} + x_{ce} \equiv x_{ba} + x_{bc} + x_{da} + x_{dc}$ at increasing, quadratic cost

$$\delta(x_{ae} + x_{ce})^2 \qquad \delta > 0.$$

We assume that each tier of firms is essential, i. e. a coalition of firms can only produce some final output if it contains player $e$, plus at least one of $a$ and $c$, plus at least one of $b$ and $d$.[28] Standard computations show that the worth of the grand coalition $N = \{a, b, c, d, e\}$ is equal to

$$v(abcde) = \frac{\alpha^2}{1 + 4\delta + \gamma},$$

which is well-defined for all $\delta > 0$ and $\gamma \geq -1$. Total output is

$$x_{ba} + x_{bc} + x_{da} + x_{dc} = \frac{2\alpha}{1 + 4\delta + \gamma}.$$

Now consider the strategic effect of a vertical merger of firms $a$ and $b$ on the profitability of a second merger of $c$ and $d$. By (4.25) and the maximization of the relevant subcoalition payoffs, this is equal to

$$\begin{aligned} &-\frac{2}{120}\Delta^2_{ab}\big[\Delta^3_{cde}(\emptyset)\big] \\ =&-\frac{2}{120}\alpha^2\left(\frac{1}{1+\delta} + \frac{1}{1+4\delta+\gamma} - \frac{1}{1+2\delta} - \frac{1}{1+2\delta+\gamma}\right). \end{aligned} \tag{4.26}$$

As noted above, this is a mutual effect of each vertical merger on the profitability of the other. The right-hand side of (4.26) is a continuous, decreasing function of the substitutability parameter $\gamma$. An example is shown as the solid curve in Figure 4.3. It is positive when the goods are perfect complements ($\gamma = -1$) and negative when they are perfect substitutes ($\gamma = 1$). Thus, the two vertical-merger decisions are strategic complements if and only if the goods produced by the intermediate firms are sufficiently complementary (precisely, if $\gamma$ is lower than the unique value $\gamma_0$ at which (4.26) is zero).

[28]This assumption simplifies the analysis by reducing the worth of some subsets of $N$ to zero; it is not critical.

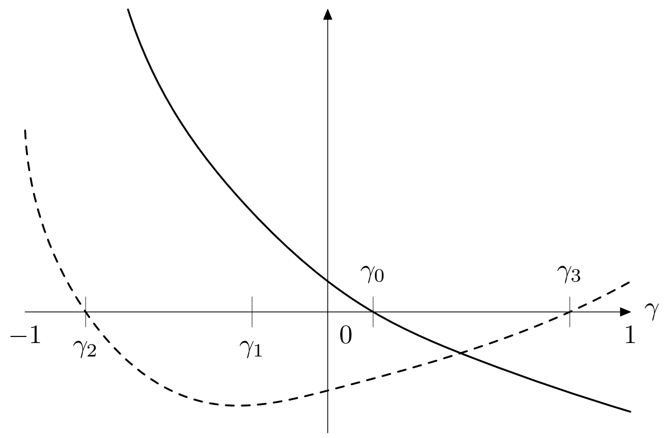


**Figure 4.3:** Profitability of a vertical merger in isolation (broken line), and mutual effect of one merger on the other (solid line). The two goods are perfect complements when $\gamma = -1$ and perfect substitutes when $\gamma = 1$. The vertical axis indicates utils. $\alpha = 2$ and $\delta = .3$

In the example drawn in Figure 4.3 with $\alpha = 2$ and $\delta = .3$, the range of substitutability parameters $\gamma$ is segmented into four cases. First, when $\gamma \in [-1, \gamma_2]$, so that the goods are close complements, $c$ and $d$ profit from merging regardless of the decision taken by the other merger candidates (but their total profit is higher if $a$ and $b$ also merge).

Second, when $\gamma \in (\gamma_2, \gamma_1]$ then the merger of $c$ and $d$ is profitable if and only if $a$ and $b$ merge as well: The broken curve is negative but the sum of both curves is positive. By the fundamental symmetry of this simple model, the incentives for $a$ and $b$ are exactly the same, and it is precisely the strategic complementarity of the two mergers which makes each of them profitable if both occur. Essentially, this leads to a simple $2 \times 2$ normal-form game with two symmetric, Pareto-ranked equilibria in which both pairs of players take the same action. It is straightforward to think of real-life situations where one of the two pairs is prevented by some exogenous force from taking the preferred merger (or de-merger) decision; when this obstacle is removed, *both* pairs or merger candidates act in unison even though only one of them is directly affected by it. Hence, even in this simple illustration can one find a purely bargaining-based rationale for a merger or de-merger "wave".

Third, when $\gamma \in (\gamma_1, \gamma_3]$, neither pair of candidates wants to merge no matter what action is taken by the other. Finally, when the goods are very close substitutes and $\gamma \in (\gamma_3, 1]$, the $c$–$d$ merger (and, by symmetry, the $a$–$b$ merger as well) is profitable if and only if the other merger does not take place. While this implies the usual kind of coordination problems, it is another instance where the reversal of either merger decision leads to a change of the other.

In the remainder of this section, we describe the forces behind these effects in more detail. To develop an intuition for the mutual effect in (4.26), it is useful first to consider the broken curve in Figure 4.3 representing the merger profitability for $c$ and $d$, $M^{cd}(cd)$, when players $a$ and $b$ remain separate. This is the negative sum of the effects of the $c$–$d$ merger on each of the remaining three players. Formally,

$$M^{cd}(cd) = -\big[M^a(cd) + M^b(cd) + M^e(cd)\big]. \tag{4.27}$$

Consider the three terms on the RHS in turn. The effect $M^a(cd)$ on player $a$ is a decreasing function of the substitutability parameter $\gamma$. Intuitively, the merger of $c$ and $d$ results in a shift towards the margin of the negotiation over the joint rent with player $a$ as a supplier to retailer $d$. The more substitutable the goods produced by $a$ and $c$, the lower is the marginal rent relative to the inframarginal one which $c$ and $d$ "lock in" by merging.[29]

In contrast, the merger benefit $M^b(cd)$ for retailer $b$ is a negative, concave and increasing function of $\gamma$. The $c$–$d$ merger shifts the negotiation between retailer $b$ and supplier $c$ to the margin where rents are lower due to the convexity of the upstream firm's costs.[30] This effect is more pronounced when output (and hence, marginal cost) is high, which is the case when the goods are complements. Conversely, total output falls in $\gamma$, so when the goods are more substitutable, marginal costs are relatively lower and the $c$–$d$ merger is less detrimental to retailer $b$.

Finally, the merger benefit $M^e(cd)$ of the upstream firm $e$ is a positive function which monotonically decreases in $\gamma$. Being an essential player, $e$ always increases the complementarity of $c$ and $d$. The more complementary the *goods* produced by the intermediate firms $a$ and $c$, the higher is the worth of the grand coalition $v(N)$ and the larger is the average contribution of the upstream firm $e$ to the complementarity of the *players* $c$ and $d$, and hence, $M^e(cd)$.

Overall, (4.27) is a quasiconvex function of $\gamma$ on $[-1, 1]$ which is decreasing

[29]This is closely related to the "scarce needs" variant of the vertical-foreclosure model of Hart and Tirole (1990).

[30]Cf. the "scarce supplies" variant in Hart and Tirole (1990).

at $\gamma = -1$. If $\delta < 1/2$, then there is an interval $[\underline{\gamma}, \overline{\gamma}]$, $-1 \leq \underline{\gamma} < \overline{\gamma} \leq 1$, on which (4.27) is negative.[31] Otherwise, it is positive throughout. Intuitively, when the goods are close complements, then the negative effect of the $c$–$d$ merger on retailer $b$ dominates the effects on the other two outsiders, so that $c$ and $d$ profit from merging. Moreover, when the cost parameter $\delta$ is large ($> 1/2$), the effect on $b$ dominates regardless of the substitutability between the goods, and the merger of $c$ and $d$ is profitable in isolation for all values of $\gamma$.

Now consider the solid curve in Figure 4.3 depicting the strategic effect of a merger of players $a$ and $b$ on the profitability of the $c$–$d$ merger. After $a$ and $b$ have merged, they employ or withdraw their joint resources together rather than separately. Thus, the merger of $a$ and $b$ eliminates those bargaining situations where the upstream firm $e$ makes its marginal contribution to a set of players containing either $a$ or $b$ but not the other. As it turns out, this implies that the merger benefit *from the $c$–$d$ merger* for the upstream firm, $M^e(cd)$, is reduced when the goods are complements and increased when they are substitutes.

To see this, consider the terms which make up $M^e(cd)$ when $a$ and $b$ are separate players. Recall that

$$M^e(cd) = \frac{1}{n!} \sum_{\pi \in \Pi} 1[c \prec_\pi e \prec_\pi d] \Delta_e\big[\Delta^2_{cd}(B'(e,\pi))\big].$$

The preceding coalitions $B'(e,\pi)$ are $\{\emptyset\}$, $\{a,b\}$, $\{a\}$, and $\{b\}$ (with different but positive weights). In the absence of the intermediate firm $a$, there is no second good, and both $\Delta_e(\Delta^2_{cd}(\emptyset))$ and $\Delta_e(\Delta^2_{cd}(b))$ are positive but independent of the substitution parameter $\gamma$. In contrast, $\gamma$ has an important effect for $\Delta_e(\Delta^2_{cd}(a))$ and $\Delta_e(\Delta^2_{cd}(ab))$.

$\Delta_e(\Delta^2_{cd}(a))$ represents a bargaining situation where retailer $b$ is not active; the term is positive and monotonically decreasing in $\gamma \in [-1, 1)$ and zero for $\gamma = 1$. With complementary goods, the essential player $e$ enables $c$ and $d$ to generate together, but not separately, a positive profit which increases as the total value of the final output increases (i. e. as $\gamma$ decreases and the goods become more complementary). In contrast, as $\gamma$ goes to unity and the goods become perfect substitutes, player $c$ no longer provides any additional value when its competitor $a$ is present, so that $c$ and $d$ are no longer complementary players.

On the other hand, $\Delta_e(\Delta^2_{cd}(ab))$ determines the bargaining share of the upstream firm $e$ when all other players are active as well; it is negative when the

[31] In Figure 4.3, this interval is given by $[\gamma_2, \gamma_3]$.

goods are close complements, i. e. the joint contribution of supplier $c$ and retailer $d$ to the remaining three players is less than the sum of their individual contributions. This is because a second retailer adds no value when the two intermediate firms $a$ and $c$ supply perfect complements produced (or procured) at increasing cost. The marginal value of an additional unit is then equal to $\alpha$ at either retailer independent of absolute quantities, so adding $d$ as a second retailer does not increase the worth of coalition $\{a, b, c, e\}$. In contrast, retailer $d$ does make a positive contribution when supplier $c$ is absent, since without the complementary good produced by $c$, the marginal value $\alpha - 2x_{ia}$ of supplier $a$'s output $x_{ia}$ is strictly decreasing in each local retailing market. Hence, supplier $c$ reduces the marginal contribution of retailer $d$ when the goods are close complements, resulting in a weak bargaining position for the essential upstream firm $e$. When the goods are perfect substitutes, in contrast, the marginal contribution of supplier $c$ is zero whenever its competitor $a$ is already present, so that $\Delta_e(\Delta^2_{cd}(ab)) = 0$ for $\gamma = 1$.

When $a$ and $b$ merge, the upstream firm $e$ bargains either in the presence of the merged entity or in the absence of both $a$ and $b$. Technically, the third-difference terms $\Delta_e(\Delta^2_{cd}(a))$ and $\Delta_e(\Delta^2_{cd}(b))$ are replaced by $\Delta_e(\Delta^2_{cd}(\emptyset))$ and $\Delta_e(\Delta^2_{cd}(ab))$. Since $\Delta_e(\Delta^2_{cd}(b))$ and $\Delta_e(\Delta^2_{cd}(\emptyset))$ are constant in $\gamma$, and since

$$\Delta_e(\Delta^2_{cd}(a)) - \Delta_e(\Delta^2_{cd}(ab))$$

is positive and monotonically decreasing in $\gamma$, the merger of $a$ and $b$ reduces the bargaining power of the upstream firm $e$ if the goods are complementary (i. e. if $\gamma$ is low). The upshot is that the mutual effect of the two disjoint mergers on each other represented by the solid curve in Figure 4.3 is a decreasing and convex function of $\gamma$ which, for $\delta > 0$, is positive at $\gamma = -1$ and negative at $\gamma = 1$. Hence, there is a unique value $\gamma_0 \in (-1, 1)$ such that the merger decisions of the two pairs of candidates are strategic complements if $\gamma < \gamma_0$ and strategic substitutes otherwise.

When each pair of merger candidates takes into account the strategic effect in (4.26) in addition to the direct effect of their decision in (4.27), then the range of values of the substitutability parameter $\gamma \in [-1, 1]$ is segmented into different cases as a function of the sum of both effects as described above. Depending on parameter values, however, some of the cases described above may not exist. For instance, when the cost parameter $\delta$ is larger than 1/2, then $M^{cd}(cd)$ (the broken curve in the graph) is positive for all admissible values of $\gamma$, eliminating the interesting case where each merger takes place if and only if the other does as well.

## 4.6 Conclusion

By explicitly allowing for the heterogeneity of outside players, we extend the cooperative approach of Segal (2000) in a direction which we find particularly relevant for the study of merger and de-merger decisions. Real-life industries with bilateral bargaining are usually more complex than the two-tier models analyzed in most of the literature. In particular, intermediate-level firms face both upstream and downstream trading partners, making the assumption of essentially homogeneous outsiders rather questionable. Even in simple two-sided industries, outsider heterogeneity can arise quite naturally, as our four-firm retailing example demonstrates. Indeed, from a bargaining perspective the salient feature of a multi-tiered industry is not the number of tiers or the direction of supply flows, but the fundamental heterogeneity of the players with respect to their marginal contribution to any given coalition.

When outsiders are heterogeneous, a profitable merger can benefit some of them even in a pure bargaining model without the monopolization effects typical for Cournot-market mergers à la Salant et al. (1983). In those models, the merging firms lose market power relative to the outsiders, but the merger reduces the coordination problem among the Cournot competitors and moves the industry closer to the cartel outcome; hence, the merging firms receive a smaller share of a larger pie, and the net effect can go in either direction. In contrast, in our purely bargaining-based model (like in Segal (2000) and Inderst and Wey (2001)), the players can coordinate on maximizing total profit regardless of the market structure, and mergers are all about bargaining power and the distribution of the fixed surplus. The novel feature in this paper is that mergers lead to a redistribution among the outsiders as well as from all outsiders to the merging firms. In particular, the main share of the redistributed surplus can be received by (some) outsiders.

The fact that outsiders can benefit from a profitable merger is of particular interest when, prior to the merger decision, an outsider $k$ receives an exogenous shock to its profitability. Besides the obvious direct payoff effect, such a shock has an impact on the merger incentives of firms other than $k$. In particular, the shock can trigger a merger which overcompensates the direct payoff effect, in which case the incidence of the shock to one player falls entirely on others who may not even be directly linked to him. Among other things, this implies inefficient incentives for (binding) technology choice and R&D effort when a firm's trading partners subsequently take a merger or de-merger decision.

Finally, we show that the existence of different kinds of outsiders is a necessary and (generically) sufficient condition for the interaction of merger decisions. If the only outsiders to a merger of two players $a$ and $b$ are the potential participants in another, disjoint, merger, then there is no effect of the first merger on the profitability of the second, and hence no interaction between merger decisions. In contrast, if there are additional outsiders who do not take part in either merger, then the two pairs of merger candidates interact symmetrically, turning merger decisions into strategic substitutes or complements. This suggests a perspective on merger waves which is derived entirely from bargaining considerations: When mergers interact, then an exogenous effect leading to a revision of one merger decision can affect another, seemingly unrelated one by other players in the same industry.

## 4.7 Appendix

### 4.7.1 Bargaining and the Shapley value

To motivate our use of the Shapley value, we generalize the stability condition for multiple bilateral negotiations introduced by Stole and Zwiebel (1996a). Under some additional assumptions, Stole–Zwiebel stability implies that individual payoffs follow the "fair allocation" rule of Myerson (1980), so by his central result they are equal to the players' Shapley values. The additional assumptions (which are sufficient but not necessary) are (*i*) "efficient" bargaining sets; (*ii*) absence of externalities; and (*iii*) a mild form of quasi-concavity of the aggregate profit function.

Myerson shows that individual payoffs in a cooperative game are equal to Shapley values if (*i*) the players are connected by chains of "coordination conferences" (trading links, in our terminology), and (*ii*) the payoff allocation rule is efficient and fair; it is efficient if individual payoffs sum to the worth of the grand coalition, and it is fair if eliminating a trading link $ij$ has the same effect on the payoffs of both players $i$ and $j$,

$$X_i(L) - X_i(L \setminus ij) = X_j(L) - X_j(L \setminus ij),$$

for all trading links $ij$, where $X_i$ and $X_j$ are payoffs and $L$ is the set of all trading links.[32]

Stole and Zwiebel's stability condition applies to a setting where three or more players bargain in bilateral negotiations. Its key feature is the assumption that contracts are non-binding, so that after the breakdown of any single negotiation all remaining contracts can be renegotiated. Thus, disagreement payoffs are given by what the players can achieve by subsequent renegotiations rather than by what is specified in their expected equilibrium contracts. More precisely, we define stability as follows.

**Definition 4 (Stole–Zwiebel stability)** *The outcome of a set of simultaneous bilateral negotiations is Stole–Zwiebel (SZ) stable if no player has an incentive to re-open any of his bilateral bargaining sessions, given the contracts specified in all other negotiations, and given that all remaining contracts can be renegotiated should the re-opened bargaining session break down in disagreement.*

Let $n$ firms $N = \{1, \ldots, n\}$ bargain simultaneously and on a bilateral basis. Any two firms $i$ and $j$ in $N$ can bargain over a pair of quantities $q_{ij} \in \mathbb{R}^2_+$ which they supply to each other, as well as a net lump-sum payment $p_{ij}$. Thus, allocation and bilateral rent-sharing can be determined separately, so that the issue of double marginalization does not arise. However, the firms cannot negotiate anything other than $q_{ij}$ and $p_{ij}$. In particular, their behaviour in the remaining negotiations is not contractible. We follow the literature in assuming that if a firm is involved in several negotiations, it sends a different agent to each of them. An agent acts faithfully to maximize the profit of his firm in his bilateral negotiation, but he cannot coordinate his actions with other agents of the same firm.[33] Let

$$Q \equiv \{q_{ij}\}_{i \neq j \in N}$$

denote the set of all bilateral-quantity vectors. Let $P$ denote the corresponding set of bilateral net payments. Two firms $i$ and $j$ are *linked in $Q$* if and only if at

[32] There are three related types of efficiency in this context. "Efficient bargaining" refers to the fact that when the bargaining set contains both quantities and payments as separate items, then the bilateral surplus can be shared freely without compromising allocative efficiency. Myerson's efficiency of the payoff rule means that no surplus is wasted when individual payoffs are determined. Finally, the overall allocation is efficient if it maximizes total surplus across all firms.

[33] The assumption that agents of the same firm cannot communicate is made, for example, by Björnerstedt and Stennek (2000). A similar notion is implied by the contingent-contracts approach in Inderst and Wey (2001).

least one of the quantities in $q_{ij}$ is positive.[34] For further reference, let $Z \subseteq Q$ denote a set of bilateral quantities exogenously set to zero, $q_{ij} \equiv 0 \Leftrightarrow q_{ij} \in Z$. By abuse of notation, we will also refer to $Z$ as a set of eliminated links $ij$ corresponding to each $q_{ij} \in Z$.

Each firm $i \in N$ has a gross profit function $f_i(Q)$, defined as its revenue from sales to customers outside of $N$ minus its own cost of production conditional on $Q$, but ignoring payments to and from its trading partners in $N$. The net profit $g_i(Q, P)$ of each firm is the sum of $f_i(Q)$ and the net payments received from its trading partners.

In each bilateral negotiation, the two firms $i$ and $j$ choose bilaterally-optimal quantities $q_{ij}^B$ to maximize their joint profits, taking the outcomes of all other bargaining encounters as given,

$$q_{ij}^B = \arg\max_{q_{ij}} g_i(q_{ij}, Q_{-ij}, P) + g_j(q_{ij}, Q_{-ij}, P),$$

where $Q_{-ij}$ is the set of quantities in all other bilateral trades (including those of $i$ and $j$ with any third party $k \notin \{i, j\}$). In contrast, the bilateral net payment $p_{ij}$ is used to share the *incremental* joint surplus equally. Under Stole–Zwiebel bargaining, this is the difference between the joint profit $g_i(Q, P) + g_j(Q, P)$ after an agreement, and the sum of payoffs after a breakdown and the subsequent renegotiation of all remaining contracts.

**Assumption 8 (No externalities)** *Holding all other elements of $Q$ constant, a change in the bilateral quantity $q_{ij}$ only affects the gross profit $f_k(Q)$ of firm $k$ if $k \in \{i, j\}$.*

In particular, this assumption rules out production externalities among suppliers, and it implies that firms do not compete with each other on any market outside of $N$.

**Lemma 26** *Conditional on any set of eliminated links $Z \subseteq Q$, the allocation maximizing total surplus across all firms is SZ-stable.*

**Proof.** From a planner's point of view, a necessary condition for the optimality of an allocation $Q^*$ is that for all active links $ij \notin Z$,

$$q_{ij}^* = \arg\max_{q_{ij}} \sum_{k \in N} f_k(q_{ij}, Q_{-ij}^*),$$

[34] In the models of buyer-supplier bargaining of which we are a aware, bilateral trade takes place in only one direction. We allow for two-way trade to emphasize that the equivalence of Shapley values and payoffs under Stole–Zwiebel bargaining generalizes to this case.

i. e. each bilateral quantity must maximize industry profits conditional on all other trades following $Q^*$. Absent externalities, this is equivalent to

$$q^*_{ij} = \arg\max_{q_{ij}} f_i(q_{ij}, Q^*_{-ij}) + f_j(q_{ij}, Q^*_{-ij}) \tag{4.28}$$

for all links $ij$.

In contrast, players $i$ and $j$ maximize their joint surplus by choosing a bilateral quantity $q_{ij}$, taking the choices for all other links as given. The bilateral joint surplus of $i$ and $j$ is maximized by

$$q^B_{ij} = \arg\max_{q_{ij}} g_i(q_{ij}, Q_{-ij}, P) + g_j(q_{ij}, Q_{-ij}, P) \tag{4.29}$$

for all trading links $ij$. That is, each bilateral quantity must maximize the sum of the two parties' *net* profits given all other quantities and payments. But conditional on $Q_{-ij}$ and $P_{-ij}$, $q_{ij}$ affects only *gross* profits $f_i(\cdot)$ and $f_j(\cdot)$. (The bilateral payment $p_{ij} \in P$ cancels out.) Thus, (4.29) is equivalent to

$$q^B_{ij} = \arg\max_{q_{ij}} f_i(q_{ij}, Q_{-ij}) + f_j(q_{ij}, Q_{-ij}).$$

Clearly, this is satisfied by the necessary condition (4.28) for the planner's optimal set of quantities $Q^*$. ■

This establishes the SZ stability of the (constrained) efficient outcome after the breakdown of any number of trading links in the industry. However, bilateral quantity setting can also lead to SZ stable but inefficient outcomes. An obvious reason is coordination failure in the presence of complementarities (or scale economies); another is capacity constraints.[35]

The following assumptions rule out inefficient equilibria of the bilateral quantity game. The first says that the planner's optimal set of bilateral quantities is unique for any given set of links which have been eliminated. In particular, this rules out perfect substitutes. The second assumption requires that when the set of quantities is inefficient, at least one pair of firms has an incentive to move its quantity closer to the efficient quantity $q^*_{ij}$.

[35]Suppose there are two buyers and a seller whose marginal cost is zero up to its capacity and infinite thereafter, let the buyers' profits be increasing in the amount of the good they receive, and suppose that it is globally efficient for each buyer to receive half the seller's capacity. In this situation, *any* allocation that gives a share of $x$ of the seller's capacity to one buyer and $1 - x$ to the other is an equilibrium in the game where bilateral quantities are chosen independently.

**Assumption 9 (Unique efficient quantities)** *After any set $Z \subseteq Q$ of links has been eliminated, total surplus $\sum_{i \in N} f_i(Q)$ is uniquely maximized by $Q_Z^*$.*

**Assumption 10** *A inefficient set of bilateral quantities $Q \neq Q_Z^*$ contains at least one link $q_{ij}$ for which $\left|q_{ij}^B - q_{Z,ij}^*\right| < \left|q_{ij} - q_{Z,ij}^*\right|$, where*

$$q_{ij}^B = \arg\max_{q_{ij}} f_i(q_{ij}, Q_{-ij}) + f_j(q_{ij}, Q_{-ij})$$

*is the bilaterally efficient quantity for firm $i$ and $j$ given $Q_{-ij}$.*

A straightforward way to satisfy the last assumption is to define $\sum_{i \in N} g_i(Q, P)$—and thus, by Assumption 8, bilateral joint profits—as continuous and strictly quasi-concave in bilateral quantities, with a maximum in $q_{ij} > 0$ conditional on any $Q_{-ij}$.[36] However, Assumption 10 is less restrictive than that, allowing for discrete quantities and for some constrained-efficient quantities $q_{Z,ij}^B$ of zero.

So far we have established sufficient conditions for a unique and overall efficient SZ-stable outcome. In each bargaining session, the disagreement points are given by breakdown payoffs which in turn depend on the payoffs reached after another link has broken down, and so on until only one active link is left. In other words, payoffs are determined recursively. Myerson's result provides a convenient shortcut which leads directly to the Shapley value.

By assumption, bilateral payments $p_{ij}$ are used to share the incremental joint surplus equally. Thus, if the outcome is SZ-stable, the set of payments $P^*$ must be such that

$$g_i(Q^*, P^*) - g_i(Q_Z^*, P_Z^*) = g_j(Q^*, P^*) - g_j(Q_Z^*, P_Z^*), \tag{4.30}$$

where $Q_Z^*$ and $P_Z^*$ are quantities and payments after a breakdown of the link between $i$ and $j$ (i. e. $Z = \{q_{ij}\}$. In words, the payment $p_{ij}^* \in P^*$ between any trading partners $i$ and $j$ results in an equal split of their joint profit under the unique efficient allocation $Q^*$, minus their payoffs under the unique constrained-efficient, renegotiated allocation $Q_Z^*$ in the event of a breakdown of their bargaining session. Condition (4.30) is nothing other than the "fair allocation" rule of Myerson (1980) for an arbitrary set of players which are connected through a series of "conferences" (bilateral bargaining sessions, in our context). As Myerson shows, the fair-allocation rule plus the fact that payoffs sum to $G(Q^*)$ implies that the payoff for each player is indeed given by his Shapley (1953) value.

[36] See, for example, Björnerstedt and Stennek (2000) or Inderst and Wey (2001).

We would like to emphasize, however, that Assumptions 8–10 are sufficient but not necessary to ensure that SZ-stable payoffs are equal to Shapley values. Broadly speaking, they ensure that the players can coordinate on the allocation maximizing their joint surplus even if they only bargain over prices and quantities. Moreover, they are by no means necessary for our results based on Shapley values in the main text.

### 4.7.2 Proofs

**Proof of Lemma 20**

$M^i(ab)$ and $M^j(ab)$ are symmetric sums of third-difference terms $\Delta^3_{iab}(S)$ and $\Delta^3_{jab}(S')$, respectively, where $S$ and $S'$ are the preceding coalitions in the relevant orderings where each player follows $a$ but precedes $b$. When (4.9) holds, then the third-difference terms of $i$ exceed those of $j$ for all preceding coalitions $S \subseteq \hat{S} \equiv N \setminus \{a,b,i,j\}$ containing neither $i$ nor $j$. The remaining preceding coalitions are obtained by the union of each $S \subseteq \hat{S}$ and $\{j\}$ for player $i$ and by $S \cup i$ for player $j$. Fix some $S^* \subseteq \hat{S}$. By definition, $\Delta^3_{iab}(S^* \cup j) \equiv \Delta^2_{ab}(S^* \cup \{i,j\}) - \Delta^2_{ab}(S^* \cup j)$ and $\Delta^3_{jab} \equiv \Delta^2_{ab}(S^* \cup \{i,j\}) - \Delta^2_{ab}(S^* \cup i)$. But $\Delta^2_{ab}(S^* \cup i) \equiv \Delta^2_{ab}(S^*) + \Delta^3_{iab}(S^*)$ is larger than $\Delta^2_{ab}(S^* \cup j) \equiv \Delta^2_{ab}(S^*) + \Delta^3_{jab}(S^*)$ by (4.9). By the symmetry of both the preceding coalitions and the probability weights of the Shapley value, for each of the terms over which $M^j(ab)$ is summed there is a corresponding and (weakly) larger term in the sum of $M^i(ab)$. ■

**Proof of Lemma 23**

Fix a subset $S \subseteq N \setminus \{a,b,c,d\}$ with cardinality $|S|$. For each "complete" outsider, $S$ can appear as an argument in (4.15) with a positive or negative sign (depending on whether $k \in S$), corresponding to the preceding coalitions $(B''(k,\pi) \cup k)$ and $(B''(k,\pi))$ respectively. The coefficient for $S$ in (4.16) is the net number, across all outsiders, of preceding coalitions equal to $S$.

Fix any $k \in N \setminus \{a,b,c,d\}$. If $k \notin S$, then $S$ corresponds to $B''(k,\pi)$ in (4.15). Thus, $a$, $c$ and all the members of $S$ must precede the outsider $k$ in an ordering $\pi$ (and all other players must follow $k$) in order for $S$ to appear, with a negative sign, in (4.15). There are $(|S|+2)!(n-|S|-3)!$ such orderings for $k$, and there are $n-|S|-4$ outsiders $k \notin S$. Thus, in the aggregate across all complete outsiders, subset $S$ appears with a negative sign

$$(n-|S|-4)(|S|+2)!(n-|S|-3)! \tag{4.31}$$

times. Conversely, for each $k \in S$, $S$ corresponds to $B''(k,\pi)\cup k$ in (4.15), i. e. $a$, $c$ and $S \setminus k$ must precede $k$ in an ordering $\pi$ so that $S$ appears in (4.15) with a positive sign. There are $(|S|+1)!(n-|S|-2)!$ such orderings, and $|S|$ outsiders in $S$, giving a total of

$$|S|(|S|+1)!(n-|S|-2)! \tag{4.32}$$

instances where $S$ appears in (4.15) with a positive sign. The Lemma follows from netting (4.31) and (4.32). ■

**Proof of Lemma 24**

$M^b(cd)$ can be evaluated over all orderings $\pi \in \Pi$ with $c \prec_\pi a \prec_\pi d$, where each $\pi$ is obtained from each $\hat{\pi}$ with $c \prec_{\hat{\pi}} b \prec_{\hat{\pi}} d$ in (4.18) by the transposition of players $a$ and $b$. As before, let $B''(b,\hat{\pi}) \equiv B'(b,\hat{\pi}) \setminus \{a,b\}$ denote the "complete" outsiders preceding $b$ in ordering $\hat{\pi}$. Consequently, $B''(a,\pi) = B''(b,\hat{\pi})$. If $b \prec_{\hat{\pi}} a$, then $B'(b,\hat{\pi}) = B''(a,\pi)$; if $a \prec_{\hat{\pi}} b$, then $B'(b,\hat{\pi}) = B''(a,\pi)\cup a$. Hence,

$$\begin{aligned} M^b(cd) = \frac{1}{n!}\sum_{\pi\in\Pi}\Big\{ & 1\Big[c \prec_\pi a \prec_\pi \{b,d\}\Big]\left(\Delta^2_{cd}[B''(a,\pi)\cup b] - \Delta^2_{cd}[B''(a,\pi)]\right) \\ & - 1\Big[\{b,c\} \prec_\pi a \prec_\pi d\Big]\left(\Delta^2_{cd}[B''(a,\pi)\cup a] - \Delta^2_{cd}[B''(a,\pi)\cup\{a,b\}]\right)\Big\}. \end{aligned}$$

Moreover, by the definition of $B''(a,\pi) \equiv B'(a,\pi)\setminus\{a,b\}$,

$$\begin{aligned} M^a(cd) = \frac{1}{n!}\sum_{\pi\in\Pi}\Big\{ & 1\Big[c \prec_\pi a \prec_\pi \{b,d\}\Big]\left(\Delta^2_{cd}[B''(a,\pi)\cup a] - \Delta^2_{cd}[B''(a,\pi)]\right) \\ & - 1\Big[\{b,c\} \prec_\pi a \prec_\pi d\Big]\left(\Delta^2_{cd}[B''(a,\pi)\cup b] - \Delta^2_{cd}[B''(a,\pi)\cup\{a,b\}]\right)\Big\} \end{aligned}$$

and

$$\begin{aligned} M^{ab}(cd) = \frac{1}{n!}\sum_{\pi\in\Pi}\Big\{ & 1\Big[c \prec_\pi a \prec_\pi \{b,d\}\Big]\left(\Delta^2_{cd}[B''(a,\pi)\cup\{a,b\}] - \Delta^2_{cd}[B''(a,\pi)]\right) \\ & - 1\Big[\{b,c\} \prec_\pi a \prec_\pi d\Big]\left(\Delta^2_{cd}[B''(a,\pi)] - \Delta^2_{cd}[B''(a,\pi)\cup\{a,b\}]\right)\Big\}. \end{aligned}$$

Finally,

$$\Delta^2_{ab}(\Delta^2_{cd}(S)) \equiv \Delta^2_{cd}(S\cup\{a,b\}) + \Delta^2_{cd}(S) - \Delta^2_{cd}(S\cup a) - \Delta^2_{cd}(S\cup b).$$

The Lemma follows directly. ■

**Proof of Lemma 25**

This is analogous to the proof of Lemma 23. The expression in (4.22) is the sum of terms of the form $\Delta^2_{ab}\Delta^2_{cd}(S)$, where $S \subseteq N \setminus \{a, b, c, d\}$ corresponds to the set of complete outsiders $B''(a, \pi)$ preceding player $a$ in ordering $\pi$. Fix a subset $S$ with cardinality $|S|$. The corresponding term $\Delta^2_{ab}(\Delta^2_{cd}(S))$ has a positive sign in each ordering where player $c$ and the members of $S$ precede player $a$ (and all others follow), which is the case in

$$(|S| + 1)!(n - |S| - 2)!$$

orderings. The sign of $\Delta^2_{ab}(\Delta^2_{cd}(S))$ is negative if the coalition preceding $a$ consists of $c$, $b$ and $S$ (and no other players); there are

$$(|S| + 2)!(n - |S| - 3)!$$

orderings of this type. Thus, the (net) coefficient corresponding to subset $S$ in (4.22) is

$$\alpha_{|S|} \equiv (n - 2|S| - 4)(|S| + 1)!(n - |S| - 3)!$$

■

# Acknowledgements

Konrad Stahl, for his support, his enthusiasm, many long discussions over biscuits and coffee, and for his role in creating the GK and ENTER.

Martin Hellwig, for many insights and very helpful suggestions. For running the GK, and for making many a seminar worth attending.

Patrick Rey, for a very friendly reception at Toulouse and for many insightful discussions.

Everyone involved in the GK, the Tuesday Theory Workshop, ENTER, and MPSE, for contributing to this unique environment. For long and helpful discussions. For making it fun.

The Deutsche Forschungsgemeinschaft, for financial support for and through the Graduiertenkolleg "Allokation auf Finanz- und Gütermärkten" and its research program on Industrial Economics and Input Markets.

Donald Knuth et al., for putting LaTeX in the public domain, creating a standard among economists who do not normally believe in free lunches.[37]

Anja, for her patience.

*Danke.*

[37]See Lerner and Tirole (2000) for an analysis of open-source software.

# Bibliography

Armstrong, Mark (1996). Network interconnection. Discussion Papers in Economics and Econometrics No. 9625, Department of Economics, University of Southampton.

Armstrong, Mark (1998). Network interconnection in telecommunications. *Economic Journal*, 108, 545–564.

Björnerstedt, Jonas, and Johan Stennek (2000). Bilateral oligopoly. Unpublished manuscript, Research Institute of Industrial Economics (IUI), Stockholm.

Bulow, Jeremy (1982). Durable-goods monopolists. *Journal of Political Economy*, 90, 314–332.

Bulow, Jeremy (1986). An economic theory of planned obsolescence. *Quarterly Journal of Economics*, 101, 729–749.

Bulow, Jeremy I., John D. Geanakoplos, and Paul D. Klemperer (1985). Multimarket oligopoly: Strategic substitutes and complements. *Journal of Political Economy*, 93, 488–511.

Carter, Michael, and Julian Wright (1999). Interconnection in network industries. *Review of Industrial Organization*, 14, 1–25.

Choi, Jay Pil (1994). Network externality, compatibility choice, and planned obsolescence. *Journal of Industrial Economics*, 42, 167–182.

Coase, Ronald H. (1972). Durability and monopoly. *Journal of Law and Economics*, 15, 143–149.

Dessein, Wouter (1999a). Network competition in nonlinear pricing. Unpublished manuscript, ECARES, Université Libre de Bruxelles.

Dessein, Wouter (1999b). Network competition with heterogeneous calling patterns. Unpublished manuscript, ECARES, Université Libre de Bruxelles.

Dessein, Wouter (2000). Network competition in nonlinear pricing. Unpublished manuscript, ECARES, Université Libre de Bruxelles.

Ellison, Glenn, and Drew Fudenberg (2000). The neo-luddite's lament: Excessive upgrades in the software industry. *RAND Journal of Economics*, 31, 253–272.

Farrell, Joseph, and Garth Saloner (1985). Standardization, compatibility, and innovation. *RAND Journal of Economics*, 16, 70–83.

Fishman, Arthur, and Rafael Rob (2000). Product innovation by a durable-good monopolist. *RAND Journal of Economics*, 31, 237–252.

Fudenberg, Drew, and Jean Tirole (1984). The fat-cat effect, the puppy-dog ploy, and the lean and hungry look. *American Economic Review*, 74, 361–366.

Fudenberg, Drew, and Jean Tirole (1998). Upgrades, tradeins, and buybacks. *RAND Journal of Economics*, 29, 235–258.

Gans, Joshua S., and Stephen P. King (1999). Using 'bill and keep' interconnect arrangements to soften network competition. Unpublished manuscript, Melbourne Business School.

Gugler, K., D. Mueller, B. Yurtoglu, and D. Zulehner (2001). The characteristics and effects of mergers: An international comparison. Unpublished manuscript, University of Vienna.

Gul, Faruk (1989). Bargaining foundations of Shapley value. *Econometrica*, 57, 85–95.

Hahn, Jong-Hee (2000). Network competition and interconnection with heterogeneous subscribers. Unpublished manuscript, Christ Church College, University of Oxford.

Hart, Oliver, and Jean Tirole (1990). Vertical integration and market foreclosure. *Brookings Papers on Economic Activity: Microeconomics*, 205–286.

Hart, Sergiu, and Andreu Mas-Colell (1996). Bargaining and value. *Econometrica*, 64, 357–380.

Horn, Henrick, and Asher Wolinsky (1988a). Bilateral monopolies and incentives for merger. *RAND Journal of Economics*, 19, 408–419.

Horn, Henrik, and Asher Wolinsky (1988b). Worker substitutability and patterns of unionisation. *Economic Journal*, 98, 484–497.

Ichiishi, Tatsuro (1993). *The Cooperative Nature of the Firm*. Cambridge, UK: Cambridge University Press.

Inderst, Roman, and Christian Wey (2000). Market structure, bargaining, and technology choice. Discussion Paper FS IV 00-12, Wissenschaftszentrum Berlin.

Inderst, Roman, and Christian Wey (2001). Negotiated contracts in bilaterally oligopolistic industries. Unpublished manuscript, University College London and Wissenschaftszentrum Berlin.

Kamien, Morton I., and Israel Zang (1990). The limits of monopolization through acquisition. *Quarterly Journal of Economics*, 105, 465–499.

Katz, Michael L., and Carl Shapiro (1985). Network externalities, competition, and compatibility. *American Economic Review*, 75, 424–440.

Katz, Michael L., and Carl Shapiro (1994). Systems competition and network effects. *Journal of Economic Perspectives*, 8, 93–115.

Laffont, Jean-Jacques, Patrick Rey, and Jean Tirole (1998a). Network competition: I. Overview and nondiscriminatory pricing. *RAND Journal of Economics*, 29, 1–37.

Laffont, Jean-Jacques, Patrick Rey, and Jean Tirole (1998b). Network competition: II. Price discrimination. *RAND Journal of Economics*, 29, 38–56.

Lerner, Josh, and Jean Tirole (2000). The simple economics of open source. Unpublished manuscript, Harvard Business School and IDEI, Toulouse.

Liebowitz, S. J., and Stephen E. Margolis (1994). Network externality: An uncommon tragedy. *Journal of Economic Perspectives*, 8, 133–150.

Myerson, Roger B. (1980). Conference structures and fair allocation rules. *International Journal of Game Theory*, 9, 169–182.

Nahm, Jae (2000). The effect of compatibility on software supply and hardware demand: Forward and backward compatibility. Unpublished manuscript, Dept. of Economics, Hong Kong University of Science and Technology.

Nilssen, Tore, and Lars Sørgard (1998). Sequential horizontal mergers. *European Economic Review*, 42, 1683–1702.

Salant, Stephen W., Sheldon Switzer, and Robert J. Reynolds (1983). Losses from horizontal merger: The effects of an exogenous change in industry structure on Cournot-Nash equilibrium. *Quarterly Journal of Economics*, 98, 185–99.

Segal, Ilya (2000). Collusion, exclusion, and inclusion in random-order bargaining. Unpublished manuscript.

Shapley, Lloyd S. (1953). A value for $n$-person games. In Harold W. Kuhn, and Albert W. Tucker (eds.), *Contributions to the Theory of Games*, Volume 2, 307–317. Princeton, NJ: Princeton University Press.

Stigler, George J. (1950). Monopoly and oligopoly by merger. *American Economic Review*, 40, 23–34.

Stole, Lars, and Jeffrey Zwiebel (1998). Mergers, employee hold-up and the scope of the firm: An intrafirm bargaining approach to mergers. Unpublished manuscript, University of Chicago, GSB, and Stanford University, GSB.

Stole, Lars A., and Jeffrey Zwiebel (1996a). Intra-firm bargaining under non-binding contracts. *Review of Economic Studies*, 63, 375–410.

Stole, Lars A., and Jeffrey Zwiebel (1996b). Organizational design and technology choice under intrafirm bargaining. *American Economic Review*, 86, 195–222.

Waldman, Michael (1993). A new perspective on planned obsolescence. *Quarterly Journal of Economics*, 108, 273–283.

Waldman, Michael (1996). Planned obsolescence and the R&D decision. *RAND Journal of Economics*, 27, 583–595.