J Regul Econ (2009) 35:246–274
DOI 10.1007/s11149-009-9087-2

ORIGINAL ARTICLE

Mobile termination and collusion, revisited

Felix Höffler

Published online: 24 January 2009

Abstract The standard model by Laffont et al. (RAND Journal of Economics, 29(1): 1–37, 1998a; 38–56, 1998b) treats termination fees as an instrument to increase market power in a one-shot game of horizontal product differentiation. A prediction (Gans and King, Economics Letters, 71: 413–420, 2001) within this framework is that, with non-linear tariffs, firms should be interested in low termination fees. This seems to be at odds with regulatory experience in many countries. We offer an alternative approach, using an infinitely repeated Bertrand competition. We focus on symmetrical calling patterns and investigate simple two-part tariffs for two customer types, as well as general non-linear tariffs for two types and for a continuum of types. In this framework, when looking also at collusion in retail prices, termination fees make deviations from the collusive outcome less attractive. The optimum deviation strategy is usually to try to attract the high valuation customers since they exhibit the highest profits. Thus, a deviator will have a pool of heavy users which will have more outgoing than incoming calls, implying net termination payments. A cooperatively chosen termination rate can increase the deviator's cost and thereby always stabilizes collusion.

Keywords Two way access · Mobile telecommunications · Non-linear tariffs

JEL Classifications L41 · L43 · L96

F. Höffler (✉)
Otto Beisheim School of Management, WHU, Burgplatz 2, 56179 Vallendar, Germany
e-mail: felix.hoeffler@whu.edu

F. Höffler
Max Planck Institute for Research on Collective Goods,
Bonn, Germany

1 Introduction

There is a long discussion on excessive mobile termination rates, at least in Europe, under the calling-party-pays principle. This has led to numerous investigations by regulators (see, e.g., Competition Commission 2002; Ofcom 2003) and finally to European legislation, forcing each regulator explicitly to investigate and regulate the market for mobile termination (European Commission 2003). That regulators need to control termination fees suggests that mobile operators have an interest in relatively high termination rates [see IRG (2002), p. 11, and IRG (2004), p. 33, for an assessment of the termination fees in Europe by the Independent Regulators Group].

This interest might be partially due to the fact that non-reciprocal termination fees between mobile operators, on the one hand, and fixed line operators, on the other, serve as a means for the mobile sector to exploit the fixed line sector.[1] However, since, at least in Europe, all telecommunications incumbents are (or, in the case of British Telecom: were) integrated players it is unlikely that this is the only explanation (integrated players could oppose, via their mobile operations, coordination on termination fees that are only targeted at exploiting their own fixed line business). Thus, there is a strong suspicion that termination fees are used as an anti-competitive device in the mobile market.

This has given rise to a large, mainly theoretical, literature on two-way access. Seminal articles are by Armstrong (1998) and Laffont et al. (1998a,b) (A-LRT). The base model developed by A-LRT has been fruitfully used to investigate many additional aspects of the problem. Reviews of this literature can be found in Laffont and Tirole (2000), Armstrong (2002), and Vogelsang (2003). A short presentation focused on the problem of collusion is found in Table 1 of Peitz et al. (2004).

So—why another (theoretical) paper on this topic? The reason is that we depart from the A-LRT framework and focus on a different mechanism for collusion. In the A-LRT framework the question of collusion is posed in the following way: Do firms have an incentive to coordinate on termination fees above marginal cost, while competing in the retail market? We take the term collusion more literally, since, if firms cooperate, they might also collude in the retail market. For instance, in December 2005, the French competition authority found the three national mobile operators guilty of collusion and imposed a fine of €534 million, equivalent to about 3.5% of annual revenue of the companies involved.[2] Thus, our question is: If firms want to collude

[1] Armstrong and Wright (2007) offer an approach that deals with mobile to mobile termination and fixed to mobile termination within one model.

[2] The allegation was, however, not directly on a price cartel but on a "Yalta on market shares" via exchange of strategically important information. However, it is difficult to imagine how to freeze market shares in this industry, where all firms serve all customer groups, without implicitly agreeing not to compete too fiercely on prices. In fact, the French competition authority also suspected price coordination; however, they found prices to be hard to compare due to the high complexity of tariffs and therefore focused on the issue of unlawful exchange of information. See the presentation of Nadine Mouy, Conseil de la concurrence, at the ACE meeting, fall 2006 in Mannheim, downloadable http://www.zew.de/en/veranstaltungen/details.php?LFDNR=560. The competition authority's decision can be found at http://www.conseil-concurrence.fr/pdf/avis/05d65.pdf with an English summary http://www.conseil-concurrence.fr/user/standard.php?id_rub=160&id_article=502.

in the retail market, can this be facilitated by excessive termination fees? Since we will look at an infinitely repeated interaction, by "facilitating" we mean that collusion will be sustainable for lower values of the discount factor (similar to, e.g., Ivaldi et al. 2003).

Our alternative approach offers an explanation why mobile carriers lobby for high termination fees, while much of the existing literature rather comes to the opposite conclusion. In the A-LRT framework, and a couple of extensions (though not in all, see, e.g., Valetti and Cambini 2005; Carter and Wright 2003 or Gans and King 2001), firms cannot gain from cooperatively chosen high termination fees if they charge non-linear tariffs in the retail market [noted already by Laffont et al. (1998a), Proposition 7, partly corrected by Gans and King (2001), and generalized for heterogeneous customers by Dessein (2004a) and Hahn (2004)]. In this framework, firms should rather prefer low termination rates than high termination rates—a prediction which is at odds at least with the European experience.[3]

Our approach argues that as long as a monopolist makes higher profits on customers with high valuations (which seems plausible and in fact occurs under standard assumptions of the single crossing property and the monotone hazard rate property), termination fees can always serve anticompetitive purposes by facilitating collusion. The basic idea is that high termination fees make competitive undercutting strategies less profitable. This might best be illustrated by an example. Consider the situation of a small mobile operator who wants to increase his market share in order to better utilize his infrastructure. The operator could make calls cheaper in order to attract additional customers. However, with lower prices, his customers will make more calls and therefore also more outgoing calls than the customers of the other operators with higher prices. Thus, the operator with the low price will have to pay net termination fees to the competitors. This discourages price cuts and deviations from a collusive price level.

We capture this idea in a simple dynamic framework of collusion. n firms play an infinite repetition of a Bertrand game in non-linear tariffs. Calls are produced at constant marginal cost. Firms have a capacity constraint such that they cannot, from one period to the next, increase capacities to serve all customers. No price discrimination between on- and off-net calls is possible. Customers are heterogeneous in their marginal valuation for outgoing calls and receive no utility from incoming calls. They exhibit uniform calling patterns, i.e., everybody calls everybody else with the same probability. Collusion in this framework means that firms charge the non-linear tariff that maximizes industry profits as long as no one deviates. After deviation, they will punish (by playing a zero profit equilibrium of the stage game) forever.

Our main finding is then that introducing termination fees above marginal cost (as long as they are not too large) will always facilitate collusion in this framework.The

Footnote 2 continued
The court of first instance (Cour d'appel de Paris) upheld the decision on December 12, 2006.

[3] Our paper also departs from a standard assumption of the A-LRT framework, namely that the underlying service is differentiated. We share with papers like DeGraba (2003) or Laffont et al. (2003) the assumption that the service is a homogenous good. However, a difference to these papers is that we do not consider the case where also the receiver perceives a benefit from the service.

reason is simple. A deviator must give customers at least the same utility as they are offered under the collusion tariff. Since the collusive offer already maximizes profits, the deviator can do no better than to copy this tariff. And since in the collusive tariff higher types provide higher profits, optimum deviation implies contracting with high types only. But high types make more than the average number of outgoing calls, implying net termination costs for the deviator when termination fees exceed the marginal cost of termination.

While this basic mechanism is intuitive, it is not trivial. In the example given above, one might propose that the deviator could, in non-linear tariffs, offer contracts with a high variable cost (which reduces calls and termination fees for the operator), which customers still accept due to low fixed fees or even subsidies. Even more, it can be shown that, as long as the deviator wants to serve the same types of customers as the cartel, termination fees do actually increase deviation profits, which makes collusion more difficult to sustain. However, this can happen only if the number of different types is small. When the number of types becomes large, and can be approximated by a continuum of types, this effect disappears. The reason is that the deviator will always want to serve a different clientele than the cartel.[4] Since he has a capacity constraint, he wants to serve only high valuation types. But high types make more calls, implying net termination payments and a negative effect on the deviator's profits.

The remainder of the paper is organized as follows. Section 2 introduces the basic model. Section 3 illustrates the basic effects in a two-type model where firms are restricted to offering only a single, two-part tariff. Section 4 shows that the main insights easily carry over to the case where firms can make arbitrary non-linear offers. Section 5 contains the main result by extending the analysis to the case of a continuum of types: While in the two-type model, termination facilitates collusion only for some distributional assumptions for the types, it is shown that this restriction is an artifact of the two-point distribution. With many types, overpriced termination can always facilitate collusion. Section 6 concludes and discusses some additional aspects that are not modeled and how they might affect our argument.

2 Model

We investigate a market for a homogenous product (outgoing mobile voice calls), q. Customers have quasilinear utility functions and gain utility only from making calls, not from receiving them. Let t denote the total payment from the customers to the firms, then the utility of a customer of type θ is given by:

[4] A similar idea is modeled in Hermalin and Katz (2006b). They investigate which customer segment an entrant would target. With high access charges the entrant wants to serve only customers with an unbalanced calling pattern, i.e., with more outgoing than incoming calls. Their paper differs in that they analyze a one shot interaction and impose a certain tariff structure (only fixed subscription fees), while we look at repeated interaction and investigate (in Sect. 5) arbitrary non-linear tariffs.

Assumption 1 Quasilinear utility

$$
\begin{aligned}
&U(\theta, q, t) = u(\theta, q) - t, \\
&u(\theta, 0) = 0 \quad \forall\theta, \\
&\exists \overline{q}(\theta) \quad \forall\theta \text{ such that } \frac{\partial u}{\partial q} \begin{cases} >0 & \text{for} \quad q < \overline{q}(\theta), \\ =0 & \text{for} \quad q = \overline{q}(\theta), \\ <0 & \text{for} \quad q > \overline{q}(\theta), \end{cases} \\
&\text{and} \quad \frac{\partial^2 u}{\partial q^2} < 0 \ \forall\theta.
\end{aligned}
$$

The reservation utility is independent of the type and normalized to zero. There is a satiation level $\overline{q}(\theta)$ for each type, such that additional phone calls add no utility but disutility.[5] Higher types value the good more. We assume the single crossing property for customers' preferences.

Assumption 2 Single Crossing Property

$$
\frac{\partial u\left(\overline{\theta}, q\right)}{\partial q} > \frac{\partial u\left(\underline{\theta}, q\right)}{\partial q} \quad \text{for } \overline{\theta} > \underline{\theta}. \tag{1}
$$

Furthermore, we assume uniform calling patterns.

Assumption 3 Uniform calling pattern: A customer making q outgoing calls will spread these calls evenly across all other customers.

This implies that customers that make more than the average number of outgoing calls will have more outgoing than incoming calls. Dessein (2004b), p. 324, reports from data supporting this assumption. He observed that, e.g., business customers ("heavy users") called residential customers ("light users") 10% more often than vice versa, implying that heavy users generate more calls than they receive.

It is also important to note that this implies that the "call balance" of any firm, i.e., the difference between outgoing and incoming calls, is always zero if one's own customers make the same number of outgoing calls as the average of the customers of all the other firms (independent of the market share of customers). We assume that the number of customers is large and without loss of generality normalize it to one.

There are $n > 2$ symmetrical firms providing mobile services. All have installed a network, and network costs are sunk. We follow a standard approach in the literature (Laffont et al. 1998a, p. 6) to assume that the marginal cost of origination and termination are constant and identical, and that there are no cost advantages of carrying a call on one network rather than on two.[6] Also the marginal cost of long distance transportation (the part between origination and termination) is assumed to exhibit

[5] The only purpose of assuming a disutility beyond the satiation level is to ensure interior solutions since we will later normalize marginal production cost to zero. The results are unaffected if we would drop this assumption and at the same time introduce positive constant marginal costs for calls.

[6] See Armstrong (1998, p. 551), who makes the same assumption that there is no cost advantage of using a single network instead of two.

constant marginal cost. For simplicity, we assume that all these constant marginal costs are zero.

Assumption 4 Constant marginal cost, normalized to zero.

Firms, however, face a capacity constraint with respect to the number of customers they can serve.

Assumption 5 Capacity constraint: Each firm can serve up to β customers, where:

$$\frac{1}{n-1} < \beta < 1. \tag{2}$$

The right inequality states that no firm can serve all customers. It is actually stronger than what is necessary for our results. We need to preclude that, from one period to another, i.e., before the other firms can react, capacities cannot be expanded such that one firm takes over the whole market. Given that pricing and marketing activities are easily observable in the market, we believe that reaction times are relatively short in the mobile market. At the same time, increasing the customer base considerably within a short period is difficult, since firms usually have transitory limitations in the access to, e.g., additional mobile handsets or the availability of customer care and sales personnel. In addition, transmission capacities of the firms are limited; in particular, for peak hours, a single firm will usually not be able to transmit all traffic of all customers on its network without network extension. A special feature of mobile networks is that in many countries firms share sites, i.e., different mobile network operators use the same radio towers. The reason is that the number of physically suitable sites is limited. As a consequence, it is very difficult to conceal the extension of the own capacity from competitors. Assuming (at least in the short term) a binding capacity constraint therefore seems sensible for mobile networks.

The left inequality of (2) assumes that $n-1$ firms can serve all customers. It is crucial to support the competitive equilibrium in the Bertrand stage game, and might be violated in some mobile markets. While in our model all firms are symmetric, this is obviously not the case in reality. Thus, in some markets, dominant firms might have more than 50% of all subscribers, such that the remaining firms together might not be able to serve the whole market. However, in Western Europe, market shares in excess of 50% only exist in Belgium, Ireland and Portugal (EU, 10th Implementation Report, Annex 3, p. 52). This assumption also implies that in any (symmetric) equilibrium, firms will carry excess capacity. Although we do not model capacity decisions, a series of papers have shown that firms can have an incentive to carry excess capacity in order to be able to support collusion later on (Benoit and Krishna 1987; Davidson and Deneckere 1990).

For this framework, we consider an infinite repetition of the following stage game. In the beginning of each stage game, firms jointly determine the termination fee $\tau \geq 0$.[7]

[7] Restricting attention to termination fees above marginal costs does not qualitatively affect our results. We comment on this in the conclusions. In practice, regulators usually want τ to be “cost based”, i.e., reflect some form of “long run incremental cost”. Thus, marginal costs tend to be the lower bound for termination in practice.

They do so in the following way: each firm announces a fee τ_i, $i = 1, \dots, n$, and the fee with the majority of votes is selected by the regulators as the compulsory reciprocal termination fee for this period (in case of a draw, the regulator randomizes). We have in mind a situation where firms, for instance by using an industry association, can jointly influence how termination fees are set by the regulator. The regulator, lacking detailed cost information but being aware of the symmetry between the firms,[8] implements the suggestion of the majority. The main purpose of this assumption is that it guarantees that an optimal (equilibrium) penal code (Abreu 1988) exists which will involve setting $\tau = 0$ to ensure the harshest possible punishment (see also footnote 14).

After the termination fee of the stage game has been set, firms compete in prices in a Bertrand fashion in the customer market. All firms simultaneously announce (non-linear) prices. Each customer chooses the tariff offer that maximizes his or her utility. This interaction is infinitely repeated. Firms discount future profits at a discount factor δ.

It is well known that collusion is stable if the net present value of staying with the cartel exceeds the net present value of deviation:

$$\begin{aligned} \Pi^{Coll} + \delta V^{Coll} &\geq \widetilde{\Pi} + \delta \widetilde{V} \\ \delta &\geq \frac{\widetilde{\Pi} - \Pi^{Coll}}{V^{Coll} - \widetilde{V}}, \end{aligned} \tag{3}$$

where Π^{Coll} is the profit of a collusion period, V^{Coll} the net present value from playing the collusion strategy, $\widetilde{\Pi}$ is the profit from the optimal deviation, and $\widetilde{V}$ is the net present value for the deviating firm in the subsequent punishment. We will say that termination fees above marginal cost facilitate collusion, if, for the harshest possible punishment, (3) holds for a larger range of δ. Note that implementing the harshest possible punishment can involve not only the use of retail tariffs but also altering the termination fee.

3 Two types and two-part tariffs

There are two types of customers. The share of high valuation types $\overline{\theta}$ equals α; the share of low valuation types $\underline{\theta}$ equals $(1 - \alpha)$. In this section, firms are restricted to setting a single two-part tariff (p, f) in each period, where p denotes the variable per minute price and f the fixed fee. Let us first determine the harshest possible punishment.

Lemma 1 *In the harshest possible punishment, firms implement $\tau = 0$, and choose $p = 0$ and $f = 0$.*

Proof If $n - 1$ firms choose $\tau_i = 0$, this will be implemented, thus, also opting for $\tau_j = 0$ is an equilibrium strategy. For $\tau = 0$, firms make neither profits nor losses

[8] In reality it might well be the case that the regulator is aware of structural cost differences between operators. This is the case in Germany, where termination fees for companies operating on 1.800 MHz are higher than those operating on 900 MHz. However, knowing that costs are different does not imply that the regulator exactly knows the cost levels. Allowing for such asymmetric network operators would not qualitatively alter our analysis.

from termination; the only way to make profits is in the retail market. Given that all $n-1$ firms offer $p=0$ and $f=0$, a deviator cannot attract any customers in the retail market without making losses, since $p=0$ maximizes the gains from trade which, under the claimed equilibrium, accrue wholly to the customer. Thus, attracting customers with p or f below zero inextricably leads to losses. Offering either p or f larger than zero would leave the deviator with zero demand. There is no other equilibrium, since in any symmetric equilibrium there are idle capacities, which a firm could use by offering $\widetilde{p}-\varepsilon$ or $\widetilde{f}-\varepsilon$ and then supplying up to capacity, which is profitable for ε small enough. □

The harshest possible punishment implies zero profits forever after deviating. This can be implemented by agreeing (enforced by the punishing firms) on termination fees equal to marginal cost of zero, and playing the Bertrand zero profit equilibrium of a stage game without termination fees.[9]

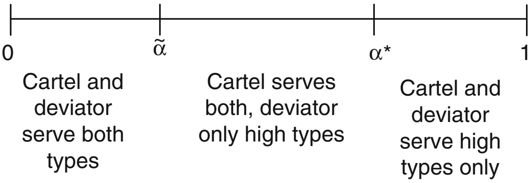
Consider a collusion equilibrium which takes this stage game as the punishment equilibrium once a firm has deviated from the collusive outcome. In the collusive outcome, either both types of customers will be served (and, since the cartel can jointly serve all customers, all customers will be served), or only the high valuation customers will be served. The latter will be the case if and only if the share of high valuation types is sufficiently large and exceeds a certain threshold. Call this threshold α^*.

Now consider an optimal deviation. Also a deviator will have to decide when making a deviating offer whether to attract both types or only the high types. It is important to observe that the threshold value of the high types (call it $\widetilde{\alpha}$) for which the deviator is indifferent between serving both types or only high types is strictly smaller than α^*. This implies that there is always a non-empty parameter region $(\widetilde{\alpha}, \alpha^*)$ where the cartel serves both types while optimal deviation (in the absence of overpriced termination) implies serving only high types.

Lemma 2 *For $\tau=0$, there exists a value $\widetilde{\alpha}<\alpha^*$ such that (i) for $0\leq\alpha\leq\widetilde{\alpha}$, the cartel serves both types and optimal deviation serves both types, (ii) for $\widetilde{\alpha}<\alpha<\alpha^*$ the cartel serves both types while optimal deviation serves only high types, and (iii) for $\alpha\geq\alpha^*$ the cartel serves both types and optimal deviation implies that both types are served.*

For extreme values of α, the cartel and the deviator behave similar: If there are almost no high types (α is very small), also low types get served. If there are almost no low types, it makes no sense to forego revenues from the high types to also attract the few low types.

[9] Laffont et al. (1998a), Proposition 7, shows that no equilibrium in the stage game with two part tariffs exists if termination fees are above cost and the transport cost is sufficiently small. In our model, transport costs are zero. However, our result differs because firms can agree on setting the termination fee equal to marginal cost in order to implement the punishment.



Different behavior of the cartel and a deviating firm

That there exists an intermediate range where the cartel and deviator behave differently is due to the capacity constraint.[10] Low types generate lower profits than high types. Thus, a firm with a capacity constraint will not want to waste its scarce capacity on low types. This is most obvious for the case that the population share of high types exceeds the capacity of a single firm, $\beta < \alpha$. Then a deviator will for sure prefer to serve only high types. If the cartel serves both types with a two part tariff, the deviator will attract the high types with a "flatter" offer. The deviator increases the fixed payment and reduces the variable payment (the optimal deviation implies a variable price equal to marginal cost of zero), which provides the same utility for the high types as the cartel offer.

We now investigate what happens when the termination fee is raised above marginal cost, $\tau > 0$. Let $q_{av} = \alpha q\left(\overline{\theta}, p\right) + (1-\alpha)\, q\left(\underline{\theta}, p\right)$ denote the average consumption per customer having accepted an offer (p, f). A deviator offering a tariff $\left(\widetilde{p}, \widetilde{f}\right)$ and serving a fraction $\widetilde{\beta} \leq \beta$ of all customers will—in the deviation period—make a profit of:

$$\widetilde{\pi} = \widetilde{\beta}\left[q_{av}\left(\widetilde{p}\right)\left(\widetilde{p} - \tau\right) + \tau\left(\left(1-\widetilde{\beta}\right) q_{av}\left(p^*\right) + \widetilde{\beta} q_{av}\left(\widetilde{p}\right)\right) + \widetilde{f}\right]. \tag{4}$$

The term in the square brackets equals the profit per customer. The first product $q_{av}\left(\widetilde{p}\right)\left(\widetilde{p} - \tau\right)$ reflects the call revenues, net of the termination fees τ. Note that this includes termination fees of "on-net calls", i.e., payments made from the deviator to himself. The second term reflects the termination revenues. Given the assumption of uniform calling patterns and a large population, these are (almost) identical for each customer, independent of his or her own calling pattern.[11] Again, these include the revenues stemming from termination payments the deviator makes to himself (the

[10] Without the capacity constraint, also for the deviator (in the deviation period) all calls might be "on-net", implying that the level of the termination fee does not matter. This point is also made in Hermalin and Katz (2006a) in their "Repeat Play" case.

[11] Note that this is approximately true only if the population is sufficiently large. Call m the number of customers. Taking into account that one does not call oneself, the exact formulation would be as follows: a $\overline{\theta}$-type receives

$$(\alpha m - 1)\frac{q\left(\overline{\theta}, p\right)}{m-1} + (1-\alpha)\, m\frac{q\left(\underline{\theta}, p\right)}{m-1}$$

incoming calls, and a $\underline{\theta}$-type receives

$$\alpha m\frac{q\left(\overline{\theta}, p\right)}{m-1} + \left((1-\alpha)\, m - 1\right)\frac{q\left(\underline{\theta}, p\right)}{m-1}$$

term $\widetilde{\beta} q_{av}(\widetilde{p})$). Finally, the deviator receives the fixed payment $\widetilde{f}$ per customer who signs the contract.

It is straightforward that whenever the deviator has the same customer base as the cartel, he will never be worse off after the introduction of overpriced termination fee. He can always mimic the collusion contract (i.e., charge the same variable price and slightly undercut the cartel's fixed fee) and, by so doing, he can sell up to capacity, which increases his profits. Since he charges the same variable price as the cartel, he avoids any net termination payments. This is due to our symmetric calling pattern assumption. The deviator can, however, do better, since the termination fees introduce a new source of revenues. By increasing the variable price p, he gets net termination payments since his customers will make less outgoing calls than the customer who stayed with the cartel.

Proposition 1 *The introduction of a termination fee $\tau > 0$ (i) does not facilitate collusion for $0 \leq \alpha \leq \widetilde{\alpha}$, and (ii) makes collusion more difficult to sustain for $\alpha \geq \alpha^*$.*

Proof See Appendix. □

For low values of α, termination fees do not affect the deviator's profits. Deviator and cartel serve both types, and a deviator will set the same two part tariff as the cartel. Matters differ for large values of α. In this case, the deviator can exploit the other firms. It sets a large price, thus its own customers have few outgoing calls compared to incoming calls. Thus, the deviator benefits from an access surplus which is high if termination is overpriced a lot. Overpriced termination will therefore make collusion less stable for high values of α.

This effect is similar to what is often discussed in the context of fixed versus mobile termination. It has been frequently argued that the mobile operators exploit the fixed line companies by setting unilaterally high termination fees (see Armstrong 2002, pp. 337–341, for a theoretical treatment).

Now consider the intermediate case. As we will argue below, for intermediate values of α, and as long as the termination fee is not too large, deviators will find it optimal to serve high types only while the cartel serves both types. In the absence of termination fees, the deviator would set $p = 0$ and extract all rent, except for the high types' information rent under the collusive outcome. This would, however, imply a non-equalized call balance. The deviator's customers would have more outgoing than incoming calls. Thus, when termination fees above marginal cost of termination are introduced, the deviator ends up with net payments for termination fees (an "access deficit" in A-LRT terminology). Therefore, introducing termination fees—as long as they are not too large—will always reduce the deviator's profits and make collusion easier to support.

Footnote 11 continued
incoming calls. Both expressions converge to $q_{av} = \alpha q\left(\overline{\theta}, p\right) + (1 - \alpha)\, q\left(\underline{\theta}, p\right)$ for $m \to \infty$, as used in (4). Note that for m small, this approximation is bad. For example, consider $m = 2$, where the first consumer makes two calls and the second zero calls. The total number of calls equals two, but the first consumer has zero incoming calls, the second two incoming calls.

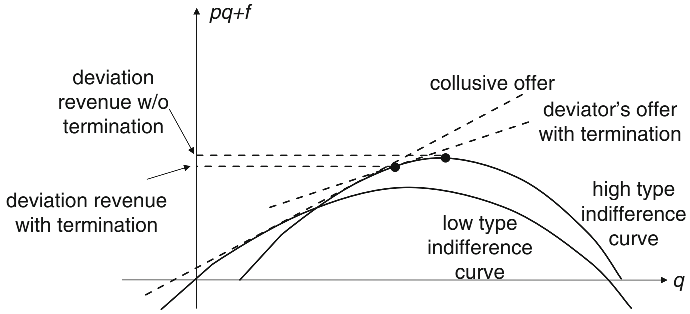


Fig. 1 Termination with two-part tariff and two types

Proposition 2 *Call* $\left(\tilde{\tilde{\alpha}}, \alpha^*\right)$ *the parameter region in which in the collusive outcome both types get served while a deviator would serve only high types. Then, for* $\alpha \in \left(\tilde{\tilde{\alpha}}, \alpha^*\right)$*, there exists a cutoff value* $\hat{\tau}$ *such that the introduction of a termination fee* $\tau \leq \hat{\tau}$ *facilitates collusion. Sufficient conditions for the interval* $\left(\tilde{\tilde{\alpha}}, \alpha^*\right)$ *to be non-empty are* $\beta < \alpha^*$ *and that* τ *is sufficiently small.*

Proof See Appendix. □

Figure 1 illustrates the effect of termination fees on the revenues from customers. Here, q denotes the quantity consumed by a certain customer. Without overpriced termination, i.e., for $\tau = 0$, a deviator—focusing on high types only—sets a price of zero and extracts all rent (except for the high type's information rent in the collusive offer) via the fixed fee. With $\tau > 0$, such a pricing would create an access deficit. Thus, the deviator chooses a positive price to reduce the high type's demand in order to save on termination fees (the slope of the offer is positive). The deviator's profits are here lower than they would be in the absence of termination fees.

This highlights the general trade-off for the deviator. He can generate income from customers or (via termination revenues) from competitors. If he sets a low usage fee, he will be able (via the fixed fee) to get a high total payment from the customers, since he increases the attainable surplus by setting the price equal to the marginal cost. With termination fees, this strategy leads to net termination payments to competitors. Setting a higher variable price reduces revenues from customers, but also payments to competitors. If the termination fee becomes very large, it will be optimal for the deviator to set a very high variable price (and possibly serve only low types) and to make his profit on the termination payments he will receive from the competitors. Therefore, only termination fees "moderately" above marginal cost can facilitate collusion.

Since termination fees allow the deviator to tap additional sources of revenues, the deviator is more tempted to serve the same customer mix as the cartel. However, if the deviator's capacity is not too large (and the termination fee τ not too big), the deviator will still find it optimal to serve high types only while the cartel serves both types, such that the effect applies that termination fees reduce the maximum deviation profit. In particular, if the deviator's capacity is smaller than the fraction of high types, the deviator will find it optimal to serve only high types. Again, the same logic applies that the deviator does not want to waste the scarce capacity on low profit types but concentrates on the higher profits attainable from high types.

If, however, for some other reason not covered in this section's model (e.g., to exploit the fixed line sector), τ is set relatively high, the collusion outcome might be affected by τ in an unexpected manner. High termination fees can be a motivation for the cartel to set the collusion price even above the monopoly price p^* that maximizes industry profits.

Proposition 3 *If both types get served in the collusion outcome and only the high types get served in the optimum deviation, there can be values of* $\tau \in \left(\underline{\tau}, \overline{\tau}\right)$ *such that collusion will be easier to sustain if the collusion offer is* $\left(\widehat{p}, \widehat{f}\right)$, $\widehat{p} > p^*$, $\widehat{f} < f^*$. *Sufficient conditions for this are (i) that the demand functions of the types are close,* $q\left(\overline{\theta}, p\right) = q\left(\underline{\theta}, p\right) + \varepsilon$, *and (ii) that they are sufficiently convex,* $\partial^2 q\left(\theta, p\right)/\partial p^2$ *is sufficiently large.*

Proof See Appendix. □

Consider a case where δ is so low that the cartel cannot sustain collusion when realizing the monopoly solution. To make collusion possible, it has to drive a wedge between the collusive payoff and the deviation payoff. We have shown before that it can do so by increasing τ above zero. Now imagine that it cannot influence τ (or τ might already be set at τ^* and collusion is not sustainable). It can never increase the collusion payoff any further, and we have already assumed maximum punishment after deviation. The only alternative left is to alter the collusion tariff such that the deviation profit decreases. Increasing p above p^* can serve this purpose. On the one hand, this clearly reduces the collusive payoff, which is bad for sustaining collusion. On the other hand, raising p lowers the number of outgoing calls of the customers of the cartel. Thus, if a deviator has a customer base which makes more calls than the customer base of the cartel, then this will increase the imbalance of outgoing and incoming calls. And, if τ is large, this hurts the deviator so much that it overcompensates for the negative effect on the collusive payoff; consequently, the difference between the collusion payoff and the deviation payoff increases.

However, if τ is too large, the optimum deviation will no longer be to serve only high types. For very high τ it will be optimal to deviate by serving only low types and to make profits mainly on the termination fees. Thus, it is only for intermediate values of τ that the effect of Proposition 3 can occur. The interval $\left(\underline{\tau}, \overline{\tau}\right)$ is non-empty if switching to serve only low types is very unattractive because this will sharply decrease payments from customers, while at the same time both types make approximately the same number of calls at the relevant price levels. This happens if both demand functions are close and are very convex.

If the effect of Proposition 3 occurs, the termination fees introduce an additional welfare problem since the collusion outcome does worse than the monopoly outcome. The cartel increases the price above the monopoly level just in order to prevent profitable deviation.

4 Two types and general non-linear tariffs

We just want to discuss briefly what happens if firms are not restricted to offering a single two-part tariff but can offer any number of two-part tariffs or even general non-linear tariffs. Termination fees can still reduce the deviation profit, and they therefore support collusion in this context.

First note that zero profits forever is again the harshest possible punishment. As in the previous section, firms opting for $\tau_i = 0$ and then setting $p = 0$ and $f = 0$ is again an equilibrium in the stage game. Profits from termination are again zero. Gaining positive payments from customers would always make them worse off since they already achieve the maximum utility from consuming the good at $p = 0$ and $f = 0$. And any firm would lose all customers when demanding positive payments since $n - 1$ firms can serve all customers.

Second, there will be values of α and β, such that in the collusive outcome, all types get served while a deviator serves only high types. At least this will be the case for $\beta \leq \alpha$. Again, the reason is that profits are higher on higher types and therefore the deviator does not want to "waste" his scarce capacity on serving the low valuation customers. That profits are higher on customers who exhibit a higher willingness to pay is not only intuitive, it is also a general property of "well behaved" adverse selection problems. Therefore, and for further use in the next section, we state this more formally in the following Lemma. Note that the results hold for any number of types as well as for a continuum of types.

Lemma 3 *Call $(t^*(\theta), q^*(\theta))$ the non-linear contract that maximizes industry profits under the assumptions 1 (concave, quasilinear utility and type independent reservation utility) and 2 (single crossing property). Then profits per customer are non-decreasing in types and strictly increasing in types if $q^*(\theta)$ strictly increases in θ.*

Proof See Appendix. □

The intuition for this result becomes clear from analyzing a discrete number of types. The proof is by contradiction. Were the profit on a type $\overline{\theta}$ lower than on the next lower type $\underline{\theta} < \overline{\theta}$, then a monopolist offering $(t^*(\theta), q^*(\theta))$ could just cancel the offer $(t^*(\overline{\theta}), q^*(\overline{\theta}))$. Due to the single crossing property and its implication that, in the optimum, the incentive compatibility constraints are downward binding, the high type would choose the low type's offer. By doing so, the monopolist could increase his profit, a contradiction to the claim that $(t^*(\theta), q^*(\theta))$ is profit maximizing.

Coming back to the case of two types, we note that, if the collusive outcome serves both types, the solution will have the standard properties: the high type will consume the optimum quantity (no distortion at the top), while the low type will have inefficiently low consumption and will receive only his reservation utility. A deviator who serves only the high types will mimic the collusive offer for the high types (minus perhaps some ε) and, therefore, end up with a customer pool having more outgoing than incoming calls. The same logic applies as before: introducing a not-too-large termination fee will reduce the deviation profit by increasing the "access deficit". Figure 2 illustrates this effect.

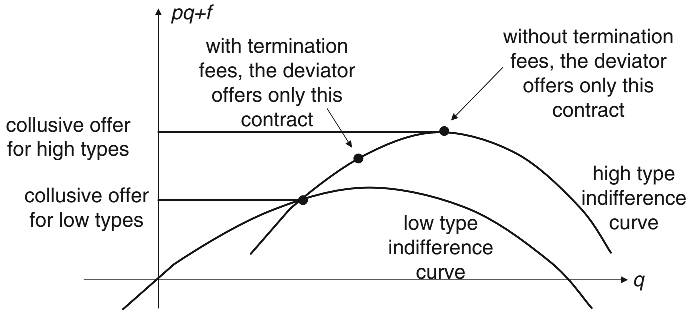


Fig. 2 Termination with general non-linear tariffs and two types

5 A continuum of types and general non-linear tariffs

Our discussion so far has shown that, under certain assumptions on the type distribution (i.e., on the parameter α), termination fees can facilitate collusion. For other parameter values of α, termination hinders collusion. In this section we want to show that this ambiguity is an artifact of the two-type case and vanishes once the number of types becomes large.

In the two-type case, termination hindered collusion if the same customer base is served in the collusion outcome and the optimum deviation. Now consider that the number of types is K. The collusion outcome serves the k highest types. When will the deviator find it optimal also to serve exactly the same k highest types? Only for very specific combinations of the distributional assumption and the capacity constraint β will this occur (types lower than k must be rare or their valuation of the service must be very low, and β must be relatively large, such that the deviator can indeed serve the k highest types). This parameter constellation will become more and more special once the number of types grows (except for the case where there are about β customers with very high valuations and the rest has very low valuations—which is again a sort of a two-type distribution). Thus, if the number of types grows, deviation will typically imply that a different customer pool is served by the deviator.

As a limit argument, we consider the case of a continuum of types, $\theta \in (0, 1)$, where—as before—firms cannot observe the customers' type. The distribution function of types is denoted by $P(\theta)$, with density $p(\theta)$. Assumptions 1, 3, 4, and 5 still hold. As before, in each stage game firms first announce τ_i and the regulator implements the termination fee which got the most votes. After that, firms compete in a Bertrand fashion by simultaneously announcing a non-linear tariff, specifying a transfer payment t for each quantity q. This stage game is infinitely repeated. In order to rely on standard solutions for the resulting adverse selection problem (see, e.g., Fudenberg and Tirole 1991, Sect. 7.3), we need some further assumptions:

Assumption 6 Monotone hazard rate property:

$$\frac{\partial}{\partial \theta}\left(\frac{p(\theta)}{1-P(\theta)}\right) \geq 0.$$

Assumption 7 Utility function: The single crossing property holds, $\frac{\partial^2 u(\theta,q)}{\partial q \partial \theta} > 0$, $\frac{\partial^3 u(\theta,q)}{\partial q \partial \theta^2} \leq 0$, and $\frac{\partial^3 u(\theta,q)}{\partial q^2 \partial \theta} \geq 0$.

To facilitate the analysis, we want to focus on the case where the collusive outcome serves all customers.[12] If firms collude, they realize the maximum profit and share this evenly as long as no firm deviates such that each firm receives Π^{Coll}. If someone deviates, they will play a punishment equilibrium. Given our assumptions, the characterization of the profit maximizing (monopolistic) solution is well known. The monopolist specifies payments and quantities such that the quantities $q^*(\theta)$ and payments $t^*(\theta)$ are strictly increasing in the types θ. By Lemma 3, this also implies that the profit per type is increasing in θ.

The harshest possible punishment again implies zero profits forever. After the deviation period, firms adjust termination fees to the marginal cost level $\tau = 0$ (again, if $n-1$ firms opt for this, and given that we assumed that a majority can implement a binding reciprocal termination fee, also opting for $\tau_i = 0$ is an equilibrium strategy). Given $\tau = 0$, and therefore termination profits of zero in any outcome, it is an equilibrium in the price game that each firm announces a linear tariff with a price equal to marginal cost. This maximizes the gains from trade for each type and implies zero profits for the firms. Firms cannot profitably deviate: attracting customers from other firms would require giving them more than the maximum surplus—which implies losses for the firm. Providing the customers with less than the maximum surplus would leave the deviating firm with no customers, since $n - 1$ firms can serve the overall market.[13]

Thus, collusion can only be supported if the payoff from colluding forever exceeds the payoff from deviating in one period and earning zero forever:

$$\frac{\Pi^{coll}}{1-\delta} \geq \Pi^{dev}.$$

Due to our assumption on the capacity constraint, no deviator can serve the whole market, but only a fraction $\beta < 1$. Again, what we want to preclude is that a firm can win the whole market from one period to the next. Therefore, if a firm deviates from the collusive outcome, it will be able to serve only a fraction of the total market. With no termination fees, $\tau = 0$, optimum deviation requires that the deviator offers a "truncated contract": he mimics the contract of the collusion outcome (possibly plus some ε benefit for the customers), but only for the highest types. The reason is again that high types generate more profit, and consequently the deviator does not want to waste his scarce capacity on low profit types.

[12] It is easy to include the case where there is a type $\underline{\theta}$ such that the cartel also serves only types $\theta > \underline{\theta}$. In this case we need to make the assumption that no single firm can serve all these customers, i.e., $\beta < 1 - P\left(\underline{\theta}\right)$.

[13] Mandy (1992) shows that there may also exist equilibria with positive profits (since our model is a model without free entry). However, if we slightly alter the assumption on the capacity constraint (2) into $1/(n-2) < \beta < 1$, i.e., $n-2$ firms can serve the whole market, then we can always construct a punishment equilibrium with zero profits as the unique outcome. Among the non-deviating firms, one (randomly selected) firm will take the role of an entrant: it will be inactive as long as there is no profitable offer possible. Then Mandy's Properties FE1–FE5 apply and a uniform price equal to the minimum of average cost (which is constant in our case) is the unique Bertrand equilibrium.

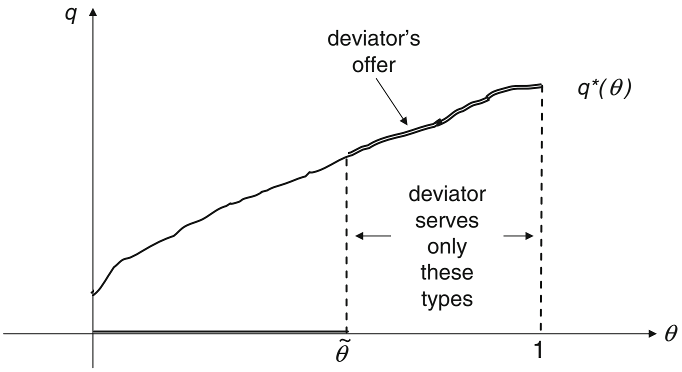


Fig. 3 Optimum non-linear tariff without termination fees

Lemma 4 *For* $\tau = 0$*, a deviator offers a truncated contract with* $q^*(\theta)$ *and* $t^*(\theta)$ *for* $\theta \geq \widetilde{\theta}$*, where* $\widetilde{\theta}$ *is defined by* $1 - P(\widetilde{\theta}) = \beta$*.*

Proof See Appendix. □

Figure 3 illustrates this Lemma. The function $q^*(\theta)$ is the solution to the monopolist's problem, i.e., the quantities for each type offered by the colluding firms. The deviator takes the quantity of the marginal customers $q\left(\widetilde{\theta}\right)$ as the minimum quantity he is willing to supply and offers the same as the collusive offer for all other types (i.e., the same price–quantity combinations (t, q) for $q \geq q\left(\widetilde{\theta}\right)$).

In the deviator's problem, the incentive compatibility constraint of $\widetilde{\theta}$ of the monopolist's problem is just replaced by $\widetilde{\theta}$'s participation constraint, i.e., $\widetilde{\theta}$ must receive at least as much utility from the deviator's offer as from the collusive outcome. Therefore, for all $\theta \geq \widetilde{\theta}$, the solutions are identical.

The result of such a strategy is that the deviator will have a customer pool that makes more outgoing calls than incoming calls (which is just a consequence of $q^*(\theta)$ being increasing). Therefore, the introduction of a not-too-large termination fee $\tau > 0$ will cause an "access deficit", i.e., net termination payments of the deviator to the other firms, and it will therefore reduce the deviation profits. This reduction in deviation profits makes collusion easier to support.

Proposition 4 *With a continuum of types and non-linear tariffs, termination fees will always facilitate collusion, as long as they are not too large.*

Proof See Appendix. □

The deviator faces a trade-off. When serving high types, he will receive high payments from the customers, but he will also have to make high termination payments to the other firms due to the imbalanced calling pattern (more outgoing than incoming calls). In order to receive at least as much profit as in the absence of termination fees, the deviator therefore would need to gain net termination income. He can do so by serving low types instead of high types, or offering contracts to high types under which they make fewer calls. Since $q^*(\theta)$ and $t^*(\theta)$ have been constructed such that they maximize the profits from any capacity $\beta, 0 < \beta < 1$, this would imply a discrete loss in payments from customers, which cannot be compensated if the termination fee is not very large. Therefore, the deviation profit will decrease when introducing a—not-too-large—termination fee.

Once the termination fee becomes too high, the deviation profit might be increasing in τ; it might even be higher than the deviation profit in the absence of termination (and thereby hinder collusion). If $\tau \to \infty$, it will obviously be optimal to serve only the lowest types and pay them not to make calls at all.

However, if firms can cooperatively set a common termination fee, they will probably avoid this. Our results rather show that they will always have an interest in setting a termination fee "moderately" in excess of marginal cost if they want to collude since this reduces the payoffs of a firm deviating from the collusive outcome.

6 Conclusion

This paper offers an explanation for why termination fees can support collusion in the retail market when mobile operators can cooperatively influence them. The key effect is that termination fees above marginal cost reduce the deviation profits. The optimum deviation strategy is usually to try to attract the high valuation customers since they are the ones with the highest profits. This strategy is made less attractive by setting termination fees above cost, since a deviator with a pool of high users will have more outgoing than incoming calls. The reduction in deviation profits stabilizes the collusion.

This is a complementary explanation to the one given by A-LRT, where termination fees can be used to "raise rivals' costs". Our approach, however, circumvents some of the problems of existence of equilibria in A-LRT and can easily show collusive effects of overpriced termination when allowing for non-linear tariffs, which is difficult in A-LRT.

In our framework, retail tariffs as well as the termination fee are part of the punishment equilibrium. While retail tariffs can be changed short term, termination fees, in particular if they are regulated, might not change quickly. Thus, there might be a time frame after the deviation period where the cartel firms can react by changing the retail tariff, but not by adjusting τ. Within this time frame, punishing the deviator is more difficult. Consider a situation where the punishing cartel firms set a two-part tariff with $p = f = 0$. Then, a deviator might set $\widetilde{p} > 0$ and $\widetilde{f} < 0$ such that he still attracts customers; however, his customers will have less outgoing calls than incoming calls, and the deviator can make termination profits which might yield positive overall profits.[14]

[14] This, however, is not an equilibrium outcome. If $\tau > 0$, we know from Laffont et al. (1998a), Proposition 7 (see footnote 8) that no equilibrium of the stage game exists. Therefore, the existence of an "optimal penal code" (Abreu 1988) is not ensured (Abreu's Assumption 2 is not ensured to hold). However, once firms can react to change τ by choosing $\tau_i = 0$, and subsequently setting $f = p = 0$, this is an equilibrium of the stage game, which already coincides with the optimal penal code, since it yields a maximum punishment of zero profits.

However, it seems unlikely that such a situation of exploitation of the cartel firms by use of overpriced termination will persist forever. The majority of firms could approach the regulator arguing that termination costs are 0 (i.e., equal to the true cost of termination in our model), and it is plausible that the regulator will react by setting $\tau = 0$. It is well understood that delaying the punishment makes deviation more attractive. However, our main argument is not affected by this since it refers to the profits of the initial deviation period, i.e., the period *before* the other firms can react at all (i.e., by changing the retail tariff): A suitably chosen termination fee $\tau > 0$ will always reduce this profit and will, ceteris paribus, make deviation less attractive.

Our analysis has abstracted from many other aspects which are highly relevant and which result in highly complex mobile tariffs.

- *Price differentiation between on- and off-net calls* Many mobile operators charge different prices for on- and off-net calls. Already Laffont et al. (1998b) discussed this issue. In our model, the tariff that maximizes industry profits does not use different prices for on- and off-net calls. If in the collusive outcome no such price differentiation is used, a deviator would make use of it. In a simple two part tariff, a deviator would charge higher off-net prices and try to attract customers by low on-net tariffs or low fixed fees. This makes collusion via termination fees harder to sustain. However, termination fees still reduce the deviator's profits since they add an additional constraint the deviator has to take care of. One might even speculate that an optimum collusive tariff might also use price differentiation to make the deviation more difficult: if the collusive offer would involve high off-net and low on-net prices, a deviation focusing on customers with high valuations would lead to an even more negative call balance.
- *Bill-and-keep* We have focused on termination fees above marginal cost. In the A-LRT framework, Gans and King (2001) find that cooperatively choosing termination fees might serve anticompetitive purposes, but will be set below marginal cost, as they are in a bill-and-keep arrangement. Such a rule makes collusion harder to sustain. Termination fees below marginal cost are equivalent to subsidizing firms with more outgoing than incoming calls. Our analysis has shown that this is to the benefit of deviators. However, for the same reason, firms that want to collude would oppose bill-and-keep arrangements.
- *Non-uniform calling patterns* Under non-uniform calling patterns, where some customers are more likely to be called than others, the basic effect of our model should still occur, as long as the calling-party-pays principle applies. In the collusive outcome, asymmetric calling patterns do not play a role because profits are shared evenly among all firms (all firms are symmetric and serve on average the same customer base). If the termination fee is not too large, a deviator would still try to focus on those customers that make many outgoing calls, since this is the primary source of revenue. Only with high termination fees, it might become interesting to attract only customers with far more incoming than outgoing calls, like call centers.

The key issue in our model is that, in the supergame, the deviator need not just mimic the collusive behavior minus epsilon. With differentiated customers, he can make higher profits per customer or per unit sold compared to the collusive outcome. While the cartel will serve "many" customers, a deviator will typically only be able to serve fewer customers and therefore will concentrate on the most attractive ones; i.e., he can engage in "cherry picking".

This might be of interest beyond its application to mobile termination, where the termination fee serves as a device to make cherry picking harder. Proposition 3 of our paper already points in an additional direction: the collusive outcome might not maximize industry profits. Rather, the cartel might need to forego profits in order to hamper cherry picking by deviators. It might try to lock in the most valuable customers (e.g., via air miles bonus programs); to achieve this, cartel members have to incur some cost (lounges, free flights). But it ties the attractive customers to the cartel for a sufficiently long period such that a deviator will not make profits before the cartel can react and punish the deviation. This type of interaction between optimum deviation and optimum collusive behavior might be of interest for future research.

Acknowledgements I am grateful for helpful comments from two anonymous referees and from Stefan Behringer, Martin Hellwig, Klaus Schmidt and Carl Christian von Weizsäcker, and seminar participants at Frankfurt, the MPI Bonn, the SFB-TR 15 conference Spring 2007, and EEA 2007. All remaining shortcomings are mine. Financial support by Deutsche Forschungsgemeinschaft through SFB-TR 15 is gratefully acknowledged.

Appendix

Proof of Lemma 2

(i) Profits on customers with high valuations are higher than on customers with low valuations: both pay the same fixed fee f, and given the uniform price and constant marginal cost, per unit profits are the same, but high valuation customers buy more units, due to the single crossing property.

(ii) Therefore, the deviator's profit is non-decreasing in α (it strictly increases if the deviator serves both types, and remains constant if the deviator serves only high valuation types).

(iii) If $\beta \leq \alpha$, the deviator will never serve low valuation types. The maximum profit per low valuation customer results from a tariff $\widetilde{p} = 0$ and $\underline{\widetilde{f}} = u\left(\underline{\theta}, \widetilde{p}\right)$, i.e., equals $\underline{\widetilde{f}} = u\left(\underline{\theta}, \widetilde{p}\right)$. By the single crossing property, this is smaller than the profit per high valuation customer in the cartel offer, $pq\left(\overline{\theta}, p\right) + f$. This again is smaller than the maximum profit the deviator can attain when serving only high valuation customers with a tariff $\widetilde{p} = 0$ and $\overline{\widetilde{f}}$, where

$$\overline{\widetilde{f}} = u\left(\overline{\theta}, q\left(\overline{\theta}, \widetilde{p}\right)\right) - u\left(\overline{\theta}, q\left(\overline{\theta}, p\right)\right) + pq\left(\overline{\theta}, p\right) + f.$$

See Fig. 4. Thus, at α^* the deviator is not indifferent but strictly prefers serving high valuation customers only. By (ii), the deviator's point of indifference

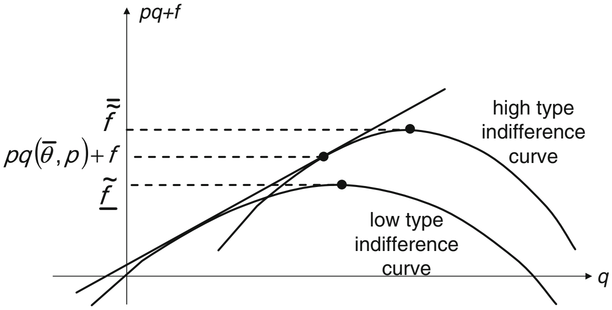


Fig. 4 Optimal type dependent deviation

between serving both types or only high valuation types, α, is strictly smaller than α^*.

(iv) If $\beta > \alpha$, the deviator might decide to set a tariff such that demand is smaller than capacity, i.e., some capacity is left idle. Due to (iii), this implies that only high valuation types are served by the deviator. At α^* the monopolist is just indifferent between serving both types and serving only high valuation types:

$$\alpha^*\pi_{hi}^* = \alpha^*\pi_{all}^{\overline{\theta}} + \left(1 - \alpha^*\right)\pi_{all}^{\underline{\theta}}, \tag{5}$$

where $\pi_{all}^{\overline{\theta}}\left(\pi_{all}^{\underline{\theta}}\right)$ denotes the profit per high (low) type customer from the profit maximizing cartel offer that serves all types. At α^*, the deviator can earn the same from serving only high valuation types (namely, $\alpha^*\pi_{\overline{\theta}}$), but strictly less when serving all types, due to the capacity constraint (namely, $\beta\left[\alpha^*\pi_{all}^{\overline{\theta}} + (1-\alpha^*)\,\pi_{all}^{\underline{\theta}}\right]$—neither can the deviator do better by altering f or p, since f^* and p^* already maximize the profit when serving both types). Then, at α^*, the deviator again strictly prefers to serve high valuation types only. And again, due to (ii), $\widetilde{\alpha}$, the point of the deviator's indifference between serving both types or only high types, must be smaller than α^*. □

Proof of Proposition 1

For (i) note that a with a deviation offer $\left(\widetilde{f} = f - \varepsilon, \widetilde{p} = p\right)$ the deviator can sell up to capacity by slightly undercutting the fixed fee while offering the same variable fee. The latter avoids any net termination payments, thus with $\tau > 0$, the deviation payoff can never be lower than the deviation profit with $\tau = 0$.

For (ii) we need to show that the deviator is indeed strictly better off by introducing termination fees for $\alpha \in [\alpha^*, 1)$, i.e., the deviator, as well as the cartel, serves high valuation customers only. If the deviator wants to serve only the high valuation types, he will maximize

$$\widetilde{\pi} = \widetilde{\beta}\left[q_{av}\left(\widetilde{p}\right)\left(\widetilde{p} - \tau\right) + \tau\left(\left(1 - \widetilde{\beta}\right)q_{av}\left(p^*\right) + \widetilde{\beta}q_{av}\left(\widetilde{p}\right)\right) + \widetilde{f}\right],$$

where

$$q_{av} = \alpha q\left(\overline{\theta}, p\right) + (1-\alpha)\, q\left(\underline{\theta}, p\right),$$

subject only to the high type's participation constraint:

$$\widetilde{f} \leq u\left(\overline{\theta}, q\left(\overline{\theta}, \widetilde{p}\right)\right) - q\left(\overline{\theta}, \widetilde{p}\right)\widetilde{p} - U\left(\overline{\theta}, p^{*}, f^{*}\right), \tag{6}$$

where $U\left(\overline{\theta}, p^{*}, f^{*}\right) \geq 0$ denotes the information rent the high valuation type receives in the collusion outcome. The deviator's pricing leads to a utilization of his capacity of $\widetilde{\beta}$, $\widetilde{\beta} \leq \beta$, i.e., the demand for the deviator's offer may be equal or below his capacity. Using (6) the deviator maximizes:

$$\begin{aligned}\max_{\widetilde{p}} \widetilde{\beta} \{ & q_{av}\left(\widetilde{p}\right)\left(\widetilde{p} - \tau\right) + \tau\left[\left(1-\widetilde{\beta}\right) q_{av}\left(p^{*}\right) + \widetilde{\beta} q_{av}\left(\widetilde{p}\right)\right] \\ & + u\left(\overline{\theta}, q\left(\overline{\theta}, \widetilde{p}\right)\right) - q\left(\overline{\theta}, \widetilde{p}\right)\widetilde{p} - U\left(\overline{\theta}, p^{*}, f^{*}\right)\}.\end{aligned}$$

Note that in this case $q_{av}(p) = q\left(\overline{\theta}, p\right)$. Using from the customers optimization that $\frac{\partial u}{\partial q} = p$, the first order conditions imply:

$$\widetilde{p} = \left(1-\widetilde{\beta}\right)\tau, \tag{7}$$

i.e., the price equals the perceived marginal cost of the deviating firm. And—after some rearrangements—profits are given by:

$$\begin{aligned}\widetilde{\pi} = \widetilde{\beta}\Big[& \underbrace{\tau\left(1-\widetilde{\beta}\right)\left[q_{av}\left(p^{*}\right) - q\left(\overline{\theta}, \widetilde{p}\right)\right]}_{\text{deviator's termination revenue}} \\ & \underbrace{+ u\left(\overline{\theta}, q\left(\overline{\theta}, \widetilde{p}\right)\right) - \underbrace{U\left(\overline{\theta}, p^{*}, f^{*}\right)}_{\text{high type's information rent}}\Big]}_{\text{payment from customer to deviator}}.\end{aligned} \tag{8}$$

Thus:

$$\begin{aligned}\frac{\partial \widetilde{\pi}}{\partial \tau} &= \widetilde{\beta}\left\{\left(1-\widetilde{\beta}\right)\left[\left[q_{av}\left(p^{*}\right) - q\left(\overline{\theta}, \widetilde{p}\right)\right] - \tau\left(1-\widetilde{\beta}\right)\frac{\partial u\left(\overline{\theta}, q\right)}{\partial q}\frac{\partial q\left(\overline{\theta}, p\right)}{\partial p}\right]\right. \\ &\quad \left. + \widetilde{p}\frac{\partial q\left(\overline{\theta}, p\right)}{\partial p}\left(1-\widetilde{\beta}\right)\right\} \\ &= \widetilde{\beta}\left(1-\widetilde{\beta}\right)\left[q_{av}\left(p^{*}\right) - q\left(\overline{\theta}, p\right)\right],\end{aligned} \tag{9}$$

which is positive, since $q_{av}(p) = q\left(\overline{\theta}, p\right)$ and $p^{*} = 0$ (from the problem of maximizing cartel profits, note that also the cartel serves only high valuation types here) and $\widetilde{p} = \left(1-\widetilde{\beta}\right)\tau > 0$ for $\tau > 0$.

Finally, if the deviator wants to serve a customer pool with strictly less outgoing calls than the average customer, this can only be optimal if the lower payment from the customers (compared to copying the collusion offer) is outweighed by termination revenues. The former is independent of the termination fee, the latter strictly increases in τ, again implying that increasing τ makes deviation more profitable. □

Proof of Proposition 2

Assume that after the introduction of a termination fee $\tau > 0$ the deviator still finds it optimal to serve high valuation customers only while the cartel serves both types of customers.

The cartel's price is determined by:

$$p^* = \frac{1}{-\frac{\partial q_{av}}{\partial p}} \left(q_{av}\left(p^*\right) - q\left(\underline{\theta}, p^*\right)\right) \tag{10}$$

$$= \frac{\alpha}{-\frac{\partial q_{av}}{\partial p}} \left(q\left(\overline{\theta}, p^*\right) - q\left(\underline{\theta}, p^*\right)\right) > 0, \tag{11}$$

where $q_{av} = \alpha q\left(\overline{\theta}, p\right) + (1-\alpha)\, q\left(\underline{\theta}, p\right)$.

From (7), the deviator's price equals $\widetilde{p} = \tau(1-\beta)$. Therefore, for τ sufficiently small, $\widetilde{p} < p^*$, implying $q_{av}\left(p^*\right) < q(\overline{\theta}, \widetilde{p})$, which by (9) implies $\frac{\partial \widetilde{\pi}}{\partial \tau} < 0$.

For τ small, profits on high valuation types are still higher than profits on low types. Thus, for $\beta < \alpha^*$, there exists an interval $\left(\widetilde{\widetilde{\alpha}}, \alpha^*\right)$, such that for $\beta \in \left(\widetilde{\widetilde{\alpha}}, \alpha^*\right)$, the same arguments (i–iii) as in the proof of Lemma 2 apply. □

Proof of Proposition 3

Assume that maximizing industry profits requires serving both types of customers while optimum deviation requires serving only high valuation customers. Consider the knife-edge case, where for given τ even with the profit maximizing contract (p^*, f^*), a deviator earns just as much from deviating as from not deviating:

$$\Pi^{Coll}(\delta) = \Pi^{Dev}(\tau) \tag{12}$$

$$\frac{1}{1-\delta}\frac{\widetilde{\beta}}{n}\pi\left(p^*\right) = \widetilde{\beta}[\tau\left(1-\widetilde{\beta}\right)\left[q_{av}\left(p^*\right) - q\left(\overline{\theta}, \widetilde{p}\right)\right] + u\left(\overline{\theta}, q\left(\overline{\theta}, \widetilde{p}\right)\right) - U\left(\overline{\theta}, p^*, f^*\right)]. \tag{13}$$

$U\left(\overline{\theta}, p^*, f^*\right)$ is the information rent, defined by:

$$U\left(\overline{\theta}, p^*, f^*\right) = u\left(\overline{\theta}, q\left(\overline{\theta}, p^*\right)\right) - p^* q\left(\overline{\theta}, p^*\right) - f^*,$$

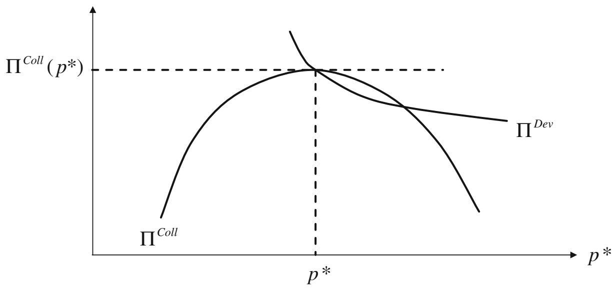


Fig. 5 Wedge between deviation profit and collusion profit

where

$$f^* = u\left(\underline{\theta}, q\left(\underline{\theta}, p^*\right)\right) - p^* q\left(\underline{\theta}, p^*\right).$$

Now consider a marginal increase in p^*. We want to show that constellations exist such that a situation as in Fig. 5 can arise: An increase of p beyond the profit maximizing level p^* can introduce a difference between the collusion payoff and the deviation payoff such that there exists an interval to the right of p^* such that collusion is easier to sustain for higher prices.

The derivative of the left hand side of (13) is zero, by construction of p^* : $\frac{\partial \pi(p^*)}{\partial p} = 0$. For the derivative of the right hand side, note that the optimum deviation price $\widetilde{p}$ is independent of p^*, see (7). Thus, the derivative of the right hand side equals:

$$\widetilde{\beta}\left\{\left[\frac{\partial q_{av}(p^*)}{\partial p}\tau(1-\beta)\right] - \frac{\partial U\left(\overline{\theta}, p^*, f^*\right)}{\partial p}\right\}. \tag{14}$$

Since the customer's optimization yields $\frac{\partial u}{\partial q} = p$, we find:

$$\frac{\partial f^*}{\partial p} = -q\left(\underline{\theta}, p^*\right),$$

and therefore,

$$\frac{\partial U\left(\overline{\theta}, p^*, f^*\right)}{\partial p} = -\left[q\left(\overline{\theta}, p^*\right) - q\left(\underline{\theta}, p^*\right)\right] < 0.$$

Thus, (14) will be negative—and the deviation profit decreasing in p—if:

$$\widetilde{\beta}\left\{[(1-\beta)]\tau\left[\alpha\frac{\partial q\left(\overline{\theta}, p\right)}{\partial p} + (1-\alpha)\frac{\partial q\left(\underline{\theta}, p\right)}{\partial p}\right] + \left[q\left(\overline{\theta}, p^*\right) - q\left(\underline{\theta}, p^*\right)\right]\right\} < 0$$

$$\tau > \frac{q\left(\overline{\theta}, p^*\right) - q\left(\underline{\theta}, p^*\right)}{(1-\beta)\left[-\alpha \frac{\partial q\left(\overline{\theta}, p\right)}{\partial p} - (1-\alpha)\frac{\partial q\left(\underline{\theta}, p\right)}{\partial p}\right]} =: \underline{\tau}, \tag{15}$$

which will be satisfied for τ sufficiently large.

However, for large values of τ it might become optimal for the deviator to serve only low valuation customers. Thus, we have to check that values of τ exist such that (15) holds and it is still optimal for the deviator to serve only high valuation types. If $\beta \leq \min(\alpha, 1-\alpha)$, the deviator will either serve only low or only high types. If he serves high types only, his profit $\overline{\pi}$ is given by:

$$\begin{aligned} \overline{\pi} = \beta \left[\tau (1-\beta) \left[\alpha q\left(\overline{\theta}, p^*\right) + (1-\alpha) q\left(\underline{\theta}, p^*\right) - q\left(\overline{\theta}, \widetilde{p}\right)\right] + u\left(\overline{\theta}, \overline{q}\right) \right. \\ \left. - U\left(\overline{\theta}, p^*, f^*\right)\right]. \end{aligned} \tag{16}$$

Serving low types only yields $\underline{\pi}$:

$$\underline{\pi} = \beta \left[\tau (1-\beta) \left[\alpha q\left(\overline{\theta}, p^*\right) + (1-\alpha) q\left(\underline{\theta}, p^*\right) - q\left(\overline{\theta}, \widetilde{p}\right)\right] + u\left(\underline{\theta}, \underline{q}\right)\right]. \tag{17}$$

Serving high types only therefore is optimal if (16 > 17):

$$\tau < \frac{u\left(\overline{\theta}, q\left(\overline{\theta}, p^*\right)\right) - u\left(\underline{\theta}, q\left(\underline{\theta}, p^*\right)\right) - U\left(\overline{\theta}, p^*, f^*\right)}{\left(q\left(\overline{\theta}, p^*\right) - q\left(\underline{\theta}, p^*\right)\right)(1-\beta)} =: \overline{\tau}. \tag{18}$$

The interval $\left(\underline{\tau}, \overline{\tau}\right)$ is non-empty if:

$$\frac{q\left(\overline{\theta}, p^*\right) - q\left(\underline{\theta}, p^*\right)}{-\alpha \frac{\partial q\left(\overline{\theta}, p\right)}{\partial p} - (1-\alpha)\frac{\partial q\left(\underline{\theta}, p\right)}{\partial p}} < \frac{u\left(\overline{\theta}, q\left(\overline{\theta}, p^*\right)\right) - u\left(\underline{\theta}, q\left(\underline{\theta}, p^*\right)\right) - U\left(\overline{\theta}, p^*, f^*\right)}{\left(q\left(\overline{\theta}, p^*\right) - q\left(\underline{\theta}, p^*\right)\right)(1-\beta)}, \tag{19}$$

which is true for any values of α for (i) $\left|\frac{\partial q\left(\overline{\theta}, p\right)}{\partial p}\right|$ and $\left|\frac{\partial q\left(\underline{\theta}, p\right)}{\partial p}\right|$ sufficiently large, (ii) $u\left(\overline{\theta}, \overline{q}\right) - u\left(\underline{\theta}, \underline{q}\right)$ sufficiently large, and (iii) $q\left(\overline{\theta}, p^*\right) - q\left(\underline{\theta}, p^*\right)$ small. This can be satisfied for $q\left(\overline{\theta}, p^*\right) = q\left(\underline{\theta}, p^*\right) + \varepsilon$ and $q\left(\underline{\theta}, p\right)$ sufficiently convex in p. See Fig. 6 for an illustration, where $u\left(\overline{\theta}, \overline{q}\right) - u\left(\underline{\theta}, \underline{q}\right)$ is the shaded area, which can be arbitrarily large. □

Proof of Lemma 3

(i) *Discrete case* Call $(q^*(\theta), t^*(\theta))$ the profit maximizing contract. It is well known that in the optimum, the incentive compatibility constraints are "down-

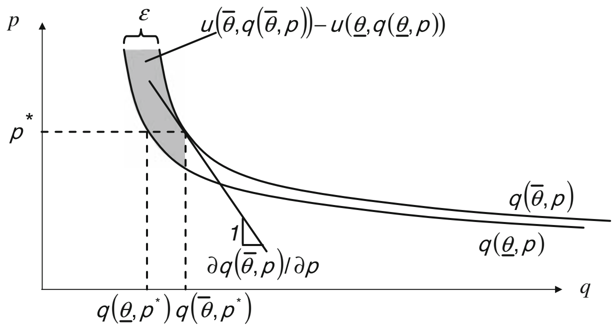


Fig. 6 Illustration of (19)

ward binding", i.e.:

$$u\left(\overline{\theta}, q^*\left(\overline{\theta}\right)\right) - t^*\left(\overline{\theta}\right) = u\left(\overline{\theta}, q^*\left(\underline{\theta}\right)\right) - t^*\left(\underline{\theta}\right). \tag{20}$$

Thus, if the contract would not include the offer $\left(q^*\left(\overline{\theta}\right), t^*\left(\overline{\theta}\right)\right)$, type $\overline{\theta}$ would choose $\left(q^*\left(\underline{\theta}\right), t^*\left(\underline{\theta}\right)\right)$. If this would increase profits, $(q^*(\theta), t^*(\theta))$ would not be profit maximizing, a contradiction.

(ii) *Continuum case* For the proof we can rely on well known characteristics of the solution to the adverse selection problem with a continuum of types, as presented, e.g., in Fudenberg and Tirole (1991), Sect. 7.3. For the contract $(q^*(\theta), t^*(\theta))$ to be implementable, we know that $q^*(\theta)$ must be non-decreasing, see Fudenberg and Tirole (1991), Theorem 7.2. The profit on a type θ, denoted by $\pi(\theta)$, is given by (see Fudenberg and Tirole 1991, pp. 264–265):[15]

$$\pi(\theta) = u(\theta, q(\theta)) - \int_0^{\theta} \frac{\partial u\left(q\left(\widehat{\theta}\right), \widehat{\theta}\right)}{\partial \widehat{\theta}} d\widehat{\theta}. \tag{21}$$

Differentiating (21) with respect to the type yields:

$$\frac{\partial \pi(\theta)}{\partial \theta} = \frac{\partial u}{\partial \theta} + \frac{\partial u}{\partial q}\frac{\partial q}{\partial \theta} - \frac{\partial u}{\partial \theta} = \frac{\partial u}{\partial q}\frac{\partial q}{\partial \theta}.$$

The solution to the relaxed optimization problem of the monopolist (i.e., neglecting the non-monotonicity constraint $\frac{\partial q}{\partial \theta} \geq 0$) is given by (see

[15] The exposition in Fudenberg and Tirole (1991) is in terms of a procurement problem. Thus, it needs some re-interpretation to apply it to our monopolistic seller problem. Their transfer payment $t(\theta)$ is from the principal to the agent, our $t(\theta)$ goes into the opposite direction. Their V_0 is the principal's valuation for the good, which is zero in our case, implying that the principal's utility is just equal to the transfer in our application. In their application, V_1 is the agent's valuation, which is usually interpreted to be negative (effort cost), while in our application it is positive and equal to $u(\theta, q(\theta))$.

Fudenberg and Tirole 1991, Eq. 7.12):[16]

$$\frac{\partial u}{\partial q}=\frac{1-P\left(\theta\right)}{p\left(\theta\right)}\cdot\frac{\partial^{2}u\left(q\left(\theta\right),\theta\right)}{\partial q\partial\theta}>0,\tag{22}$$

due to Assumption 7, where $P\left(\theta\right)$ denotes the cumulative density function of types and $p\left(\theta\right)$ denotes the density function. Thus, if the non-monotonicity constraint is not binding, i.e., $\frac{\partial q}{\partial\theta}>0$, profits are strictly increasing in types, $\frac{\partial\pi(\theta)}{\partial\theta}>0$. If $\frac{\partial q}{\partial\theta}=0$ for some parameter region $\left[\underline{\theta},\overline{\theta}\right]$, i.e., "bunching" occurs, profits will be the same for all types bunched under the same contract. □

Proof of Lemma 4

(i) Due to the capacity constraint, the deviator serves high valuation customers only since profits are higher for higher types, as shown in Lemma 3. (ii) The deviator makes the same offer as the monopolist for the types he wants to serve: Generally, the optimum contract when serving all types is given by Fudenberg and Tirole (1991, p. 265):

$$\frac{\partial V\left(q\left(\theta\right),\theta\right)}{\partial q}+\frac{\partial u\left(q\left(\theta\right),\theta\right)}{\partial q}=\frac{1-P\left(\theta\right)}{p\left(\theta\right)}\frac{\partial^{2}u\left(q\left(\theta\right),\theta\right)}{\partial q\partial\theta},\tag{23}$$

and[17]

$$t\left(\theta\right)=u\left(q\left(\theta,\theta\right)\right)-\underbrace{\int\limits_{0}^{\theta}\frac{\partial u\left(u\left(q\left(\tau\right),\tau\right)\right)}{\partial\tau}d\tau}_{I(\theta)},\tag{24}$$

where $V\left(\right)$ denotes the principal's valuation of the product (in our application, this is just equal to the production cost, which we normalized to zero) and $I\left(\theta\right)$ is the information rent of type θ. (23) defines the function $q^{*}(\theta)$, while (24) defines the profit maximizing payments $t^{*}(\theta)$ for each type. Each type θ, except the lowest, gets an information rent $I\left(\theta\right)>0$, which is increasing in θ. In the solution to the deviator's problem, the marginal customer $\widetilde{\theta}$ must receive $I\left(\widetilde{\theta}\right)$, as defined by (24), in order to accept the deviator's offer. Substituting $I\left(\widetilde{\theta}\right)$ into (24) shows that the optimum quantities $q^{*}(\theta)$ and optimum payments $t^{*}(\theta)$ for $\theta\geq\widetilde{\theta}$, do not depend on values of $\theta<\widetilde{\theta}$, and they therefore coincide with the monopolistic solution for these types. □

[16] Note that their expression $\partial V_0/\partial x$ is zero in our application, see the footnote above. Their $\partial V_1/\partial x$ is our $\partial u/\partial q$.

[17] Note that in Fudenberg and Tirole's exposition, the transfer payment is from the principal to the agent, while in our application it is in the opposite direction, explaining the different sign in our exposition.

Proof of Proposition 4

The deviator's payoff has two elements, the payments from customers, which we want to call Π_C^{dev}, and the termination revenues, which we call T, which is the sum of termination revenues the deviator receives, minus the sum of termination fees he pays:

$$\Pi_\tau^{dev} = \Pi_C^{dev} + T. \tag{25}$$

Denote by Π^{dev} the maximum deviation profit in the absence of termination fees ($\tau = 0$; the outcome of Lemma 4). We need to distinguish two cases to prove the proposition.

$T < 0$: $T < 0 \Longrightarrow \Pi_\tau^{dev} < \Pi^{dev}$, which follows trivially from the fact that the deviator can never receive more from the customers than Π^{dev}. But if the deviator offers the same contract to the customers as the collusive contract, his termination revenues are strictly negative, since

$$T\left(\widetilde{\theta}, \tau\right) = \tau\left(1 - P\left(\widetilde{\theta}\right)\right)\left[\underbrace{\int_0^1 q(\theta)\, p(\theta)\, d\theta}_{\text{average incoming calls}} - \underbrace{\frac{1}{1 - P\left(\widetilde{\theta}\right)} \int_{\widetilde{\theta}}^1 q(\theta)\, p(\theta)\, d\theta}_{\text{average outgoing calls}}\right] < 0, \tag{26}$$

since $q(\theta)$ is increasing.

$T \geq 0$: Consider a contract $\left(\widetilde{q}(\theta), \widetilde{t}(\theta)\right)$ that results in an equalized call balance for the deviator, i.e., the deviator has on average the same number of in- and outgoing calls per customer:

$$\left(\widetilde{q}(\theta), \widetilde{t}(\theta)\right) \rightarrow T = 0 \quad \forall \tau > 0.$$

Denote the resulting payments from customers, which, in this case, equal the deviator's total profit, by $\Pi\left(\widetilde{q}(\theta), \widetilde{t}(\theta)\right)$. This is strictly lower than the profit of a deviator in the absence of termination fees, i.e., for $\tau = 0$:

$$\Pi\left(\widetilde{q}(\theta), \widetilde{t}(\theta)\right) < \Pi^{dev}.$$

In order to achieve an equalized call balance, the deviator either (i) serves only customers with high valuations, $\theta \geq \widetilde{\theta}$, but with a lower quantity, or (ii) serves also some low valuation customers, $\theta < \widetilde{\theta}$. In either case, this reduces profits discretely and independent of the size of the termination fee τ. For (i), customers will accept a lower

$\widetilde{q}(\theta)$ only if also $\widetilde{t}(\theta)$ is strictly lower. Note that the decrease in quantity cannot be only marginal to achieve an equalized call balance, since the term in the square brackets in (26) is strictly negative. For (ii), the deviator must substitute some high types by low types in his customer pool. Even if he extracts the maximum profit from these types, this reduces profits compared to the case without termination fees, since, by Lemma 3, profits on lower types are lower.

By the same argument, any contract that will yield positive termination revenues for the deviator, $T > 0$, will imply that the revenues from the customers will be even smaller than $\Pi\left(\widetilde{q}(\theta), \widetilde{t}(\theta)\right)$:

$$\text{For } \tau > 0 : T > 0 \rightarrow \Pi_C^{dev} < \Pi\left(\widetilde{q}\,(\theta)\,, \widetilde{t}\,(\theta)\right).$$

Again, the deviator must induce his customer pool to consume less. This he can achieve only by (i) lower $\widetilde{\widetilde{t}}\,(\theta) < \widetilde{t}\,(\theta)$ or (ii) by serving lower types.

Therefore, for all values of $\tau > 0$,

$$\Pi_C^{dev} = \Pi^{dev} - \Delta, \quad \Delta > 0. \tag{27}$$

Since T approaches zero for $\tau \rightarrow 0$, there exists a cutoff value $\widehat{\tau} > 0$, such that for $\tau < \widehat{\tau}$ we have that $T(\tau) < \Delta \rightarrow \Pi_\tau^{dev} = \Pi_C^{dev} + T < \Pi^{dev}$. Since the deviation profit is smaller for these low values of τ compared to $\tau = 0$, and the punishment payoff for all future periods is left unaltered, deviation will now be no longer profitable for some low values of the discount factor δ, and thereby facilitate collusion, as stated in the Proposition. □

References

Abreu, D. (1988). On the theory of infinitely repeated games with discounting. *Econometrica, 56*(2), 383–396.

Armstrong, M. (1998). Network interconnection in telecommunications. *Economic Journal, 108*, 545–564.

Armstrong, M. (2002). The theory of access pricing and interconncetion. In M. E. Cave, S. K. Majumdar, & I. Vogelsang (Eds.), *Handbook of telecommunications economics* (pp. 295–384). Amsterdam: Elsevier.

Armstrong, M., & Wright, J. (2007). *Mobile call termination*. Working Paper, University College London.

Benoit, J.-P., & Krishna, V. (1987). Dynamic duopoly: Prices and quantities. *Review of Economic Studies, 54*, 23–35.

Carter, M., & Wright, J. (2003). Asymmetric network interconnection. *Review of Industrial Organization, 22*, 27–46.

Competition Commission. (2002). Mobile termination charges. Reports on references under section 13 of the Telecommunications Act 1984 on the charges made by Vodafone, O2, Orange and T-Mobile for terminating calls from fixed and mobile networks.

Davidson, C., & Deneckere, R. (1990). Excess capacity and collusion. *International Economic Review, 31*(3), 521–541.

DeGraba, P. (2003). Efficient intercarrier compensation for competing networks when customers share the value of a call. *Journal of Economics and Management Strategy, 12*(2), 207–230.

Dessein, W. (2004a). Network competition in nonlinear pricing. *RAND Journal of Economics, 34*(4), 593–611.

Dessein, W. (2004b). Network competition with hetorogenous customers and calling patterns. *Information Economics and Policy, 16*, 323–245.

European Commission. (2003). Commission recommendation of 11 february 2003 on relevant product and service markets within the electronic communications sector susceptible to Ex Ante Regulation in

accordance with directive (2002/21/EC) of the European Parliament and of the Council on a Common Regulatory Framework for Electronic Communication Networks and Services (2003/311/EC). Official Journal of the European Union L114/45 (8.5.2003).
Fudenberg, D., & Tirole, J. (1991). *Game theory*. Cambridge, MA: MIT Press.
Gans, J. S., & King, S. P. (2001). Using 'Bill and Keep' interconnect arrangements to soften network competition. *Economics Letters, 71*, 413–420.
Hahn, J.-H. (2004). Network competition and interconnection with heterogenous subscribers. *International Journal of Industrial Organization, 22*, 611–631.
Hermalin, B., & Katz, M. L. (2006a). Customer or complementor? Intercarrier compensation with two-sided benefits. University of California: Berkeley (mimeo).
Hermalin, B., & Katz, M. L. (2006b). Intercarrier compensation with all-you-can-eat retail pricing. University of California: Berkeley (mimeo).
IRG. (2002). Market definition for mobile termination. Independent Regulators Group Working Paper.
IRG. (2004). Market data analysis 2004 report.
Ivaldi, M., Jullien, B., Rey, P., Seabright, P., & Tirole, J. (2003). The economics of tacit collusion. Final report for DG Competition, European Commission.
Laffont, J.-J., & Tirole, J. (2000). *Competition in telecommunications*. Cambridge, MA: MIT Press.
Laffont, J.-J., Marcus, S., Rey, P., & Tirole, J. (2003). Internet interconnection and the off-net-cost pricing principle. *RAND Journal of Economics, 34*(2), 370–390.
Laffont, J.-J., Rey, P., & Tirole, J. (1998a). Network competition: I. Overview and nondiscriminatory pricing. *RAND Journal of Economics, 29*(1), 1–37.
Laffont, J.-J., Rey, P., & Tirole, J. (1998b). Network competition: II. Price discrimination. *RAND Journal of Economics, 29*(1), 38–56.
Mandy, D. M. (1992). Nonuniform bertrand competition. *Econometrica, 60*(6), 1293–1330.
Ofcom. (2003). Ofcom decision on mobile termination. http://www.ofcom.org.uk/consultations/past/wmvct/?a=87101.
Peitz, M., Valetti, T. M., & Wright, J. (2004). Competition in telecommunications: An introduction. *Information Economics and Policy, 16*, 315–321.
Valetti, T. M., & Cambini, C. (2005). Investments and network competition. *RAND Journal of Economics, 36*(2), 446–467.
Vogelsang, I. (2003). Price regulation of access to telecommunications networks. *Journal of Economic Literature, 41*(3), 830–862.