Mobile Call Termination in the UK*

Mark Armstrong† and Julian Wright‡

March 2007

Abstract

We discuss policy towards mobile call termination, illustrated by the 2002 Competition Commission enquiry into the UK mobile market. We present a model of the mobile market which includes both fixed-to-mobile and mobile-to-mobile call termination. In broad terms, the former service is likely to involve monopoly pricing if left unchecked, while the latter service—if the termination charge is jointly chosen by networks—may provide the mobile sector with the means by which to relax competition. Competition is often relaxed by choosing a low mobile-to-mobile termination charge. If feasible, then, unregulated networks often wish to set different termination charges depending on whether traffic originates on the fixed or mobile network. By contrast, social optimality often requires that uniform termination charges be imposed.

1 Background

In this paper we discuss policy towards call termination on mobile telephone networks, illustrated by the 2002 Competition Commission enquiry into the UK mobile market.[1] This was an unusually complicated enquiry (the report was printed in three volumes). The case demonstrates the utility of employing stripped-down economic models to cast light on complex interactions between firms and consumers.[2]

The investigation concerned the level of call termination charges levied by four mobile networks in the UK: O_2, *Orange, T-Mobile* and *Vodafone*.[3] Call termination refers to the

*We are grateful to Stefan Behringer, John Vickers and Helen Weeds for very helpful comments.

†University College London, email mark.armstrong@ucl.ac.uk. Armstrong acted in a limited way as advisor to Oftel, then the telecommunications regulator, during the course of the enquiry discussed in this paper.

‡National University of Singapore, email jwright@nus.edu.sg

[1] See Competition Commission (2003).

[2] Surveys of the economic literature relating to mobile call termination include Armstrong (2002, sections 3 and 4) and Gans, King, and Wright (2005).

[3] At the time of the enquiry, a fifth network, *H3G*, had just entered. However, this firm used a newer "third generation" technology which was specifically excluded from the enquiry.

service where a network completes (or ‘terminates’) a call made to one of its subscribers by a subscriber on another telephone network. There are two broad kinds of call termination on mobile networks: termination of calls made from other mobile networks (termed mobile-to-mobile, or MTM, termination in the following discussion), and termination of calls made by callers on the fixed telecoms network (fixed-to-mobile, or FTM, termination).[4]

These two forms of call termination have subtly different economic characteristics, and may lead to different forms of market failure. Broadly speaking, FTM call termination is likely to involve unilateral monopoly pricing if left unchecked. The vast majority of mobile subscribers join just one mobile network, and so callers on the fixed telephone network must route calls through a subscriber’s chosen network. A mobile network therefore holds a monopoly over delivering calls to its subscribers, and has an incentive to set high charges for granting access to these subscribers.

A similar bottleneck might be expected with MTM call termination. Indeed, when networks choose the MTM termination charge unilaterally this bottleneck is present. However, because of the greater degree of symmetry between mobile networks (relative to the situation with a fixed and mobile network), MTM termination charges are often chosen by negotiation between networks. The issue here is whether mobile networks can use negotiated termination charges in order to relax competition for subscribers. For instance, MTM termination charges directly affect each network’s cost of making a MTM “off-net” call (that is, a call to a subscriber on a rival mobile network) but does not affect a network’s cost of making an “on-net” call (that is, a call between two subscribers on the same network). As such, MTM termination charges have an impact on the balance between off-net and on-net call charges, and this in turn can affect competitive conditions in the market for subscribers. As we will see in a formal model, when networks can coordinate on a choice of MTM termination charge, they will often have an incentive to choose too *low* a charge. At the time of the enquiry, MTM termination was less significant in terms of volumes, and in 2000/01 72% of the mobile industry’s termination traffic originated on the fixed network, while the remaining 28% came from rival mobile networks.[5]

At the start of 2002, the performance of the four mobile operators is summarized as follows.[6]

[4]Two of the mobile networks, *Cellnet* (the precursor to O_2) and *Vodafone*, faced an earlier enquiry into their FTM termination charges in 1998—see MMC (1999b). This earlier enquiry did not investigate MTM termination charges, nor did it investigate FTM termination charges levied by the two more recent entrants to the market (*Orange* and *T-Mobile*). The enquiry concluded that the two networks’ FTM termination charges were too high in relation to cost, and based on its recommendations, Oftel (then the telecommunications sector regulator) regulated FTM termination by means of a price cap. It was the imminent expiry of this price cap which led to the enquiry discussed in this paper.

[5]This excludes traffic originating outside the UK. See Competition Commission (2003, Table 6.8).

[6]See Competition Commission (2003, Table 2.1). At the time of the enquiry there were also a number of mobile virtual network operators (MVNOs), who provide some extra competition at the retail level. However, they were not permitted to set different call termination charges from those levied by the “host” network, and so did not play a major role in the determination of termination charges—see Competition Commission

Table 1: Industry Statistics (2002)

	Vodafone	O_2	*Orange*	*T-Mobile*
Annual turnover (£m)	3,596	2,759	3,397	2,062
Average subscriber numbers (m)	12.8	10.8	11.1	9.4
Average revenue per year per subscriber (excluding call termination)	£276	£231	£246	£202
Outgoing call minutes per year per subscriber (including on-net calls)	1,070	787	1,161	925
Incoming call minutes per year per subscriber (excluding on-net calls)	531	454	568	500

Thus, at the time of the enquiry, the four networks were reasonably symmetrically placed. Importantly, and in contrast to the situation in many other European countries, all four mobile networks are separately owned from the significant fixed networks.[7]

In the years leading up to the enquiry, the mobile market was characterised by explosive growth. The total number of subscribers in 1997 was 7 million, whereas this number was 45 million by the end of 2001. The enquiry stated, though, that "it seems likely that there will be at best only modest further growth in customer numbers".[8]

Competition Commission (2003, Table 6.10) reports that average retail revenue per minute for all four networks in the period 2000/01 was:

Table 2: Average Call Charges

	Off-net MTM	On-net MTM	FTM	mobile to fixed
pence per minute	24.9	5.9	14.6	7.1

This shows a striking degree of differential pricing between off-net and on-net calls.[9] Due in part to this price differential, the volumes of off-net and on-net calls are rather unbalanced, and Competition Commission (2003, Table 6.5) reports that the share of outgoing call minutes from mobile networks in the period 2001/2002 was:

(2003, paras. 2.32–2.33).

[7] O_2 was previously fully owned by the dominant fixed operator, *BT*. However, the two companies demerged in November 2001. See Competition Commission (2003, chapter 5) for more details on the history and ownership structure of the four networks.

[8] See Competition Commission (2003, paras. 2.24 and 2.28). In fact, at the end of 2006 the number of subscribers was approximately 62 million (see Table 4 in Ofcom (2007)), which is greater than the current official UK population.

[9] It is a complicated, and to some extent arbitrary, task to give precise estimates for the prices of the various types of calls. This is because mobile networks each offer a wide variety of tariffs, with different monthly rentals (where applicable) corresponding to different volumes of inclusive call minutes. Apparently, the numbers reported in Table 2 do not include the networks' subscription revenue. One reason for why on-net charges appear to be so much lower than off-net charges is that, in the period before the enquiry at least, networks did not typically allow off-net calls within their bundles of inclusive calls. Competition Commission (2003, figure 6.8) shows how the differential between on-net and off-net MTM calls rose steeply in

Table 3: Shares of Types of Mobile Calls

	Off-net MTM	On-net MTM	mobile to fixed
%	14.6	30.3	55.1

With equal off-net and on-net charges and four roughly symmetric networks, one might perhaps expect that off-net traffic would be approximately three times greater than on-net traffic, rather than half as much as was actually the case. As well as the price differential, there are at least two reasons why call volumes are biased towards on-net calls. First, "closed user groups", i.e., groups of subscribers who predominantly make calls within their own group, may be present. Often, such groups have their network subscription decision made centrally (e.g., by their employer's procurement office) and to a single network. To the extent these groups are widespread, this will boost the share of on-net calls in the market.[10] Second, there may be some substitution between on-net MTM calls and FTM calls. A mobile subscriber, when she is in the home or office, has a choice between calling another mobile subscriber by means of either her fixed line or mobile phone. In many cases, she will use the cheaper alternative. From Table 2, this implies she will typically make an on-net MTM call if the recipient is on the same mobile network and she is more likely to make a FTM call if the recipient is on another mobile network. This will amplify the bias towards on-net call volumes.

The Commission distinguished between the (wholesale) market for call termination on mobile networks and the (retail) market for mobile subscribers and their outbound calls. In the former case, they concluded (para. 1.4) that each mobile network "has a monopoly of call termination on its own network. This is because there are currently no practical technological means of terminating a call other than on the network of the [mobile network] to which the called party subscribes and none that seems likely to become commercially viable in the near future. There are also no ready substitutes for calling a mobile phone at the retail level, such as calling a fixed line instead."

There was some debate in the enquiry about the competitiveness of the mobile retail market. Oftel, then the industry regulator, suggested that the retail market was not yet "effectively competitive", and this was because "of the room for improvement in consumer awareness of different prices and tariffs, evidence of the existence of some barriers to customers' switching networks and poor levels of consumer information. Prices were above the level that would be found in an effectively competitive market; and Vodafone's return on capital employed (ROCE) had consistently, and substantially, exceeded the cost of capital." Nevertheless, Oftel believed that the mobile sector was "prospectively competitive".[11] The

the period 1999 to 2002. The differential has reduced somewhat since 2002, presumably due to the reduction in termination charges since that time, although it is still substantial. Figure 3.39 in Ofcom (2006a) shows that in 2006 the average off-net call charge was 11.3 ppm compared with an average on-net call charge of 4.2 ppm.

[10] See Competition Commission (2003, paras. 2.113-2.121).

[11] Competition Commission (2003, para. 2.155). However, in 2003 Oftel determined that the mobile retail

four mobile networks argued vigorously that the retail market was competitive, and that any profits from high termination charges would be competed away and passed on to their subscribers. The Competition Commission concluded (2003, para. 2.211) that "while there is intense competition among the [mobile networks] to attract and sign up subscribers to their networks [...], there is less effective competition in call origination [...], as evidenced by high margins for off-net calls, a substantial level of unused free minutes in mobile packages, and the bundling and complexity of call tariffs. All of this indicates, in our view, less than effective competition at the retail level."

Finally, we provide data about the costs and charges for mobile call termination at the time of the enquiry. A good deal of effort by the Commission and the enquiry participants was put into estimating the costs of providing call termination, and of course much depends on the details of the accounting procedure employed—for instance, the relevant time horizon, the treatment of common and joint costs, the depreciation procedure, the cost of capital, and so on. (See Competition Commission, 2003, chapter 7.) To cut a long story short, the Commission estimated that the long-run average incremental cost of call termination, in 2000/1 prices, was about 7.1 pence per minute in 2002, and was expected to fall to 4.1 pence per minute by 2006.[12] The networks' termination charges varied by the time of day and week, and Competition Commission (2003, Table 6.13) reports the charges at the time of the enquiry to be:

Table 4: Termination Charges in March 2001 (pence per minute)

	Daytime	Evening	Weekends
O_2	12.4	12.4	1.1
Orange	15.2	11.0	4.5
T-Mobile	16.8	10.8	2.5
Vodafone	13.2	9.8	4.7

Note that the networks did not set different termination charges for FTM and MTM calls. This contrasts with much of our formal analysis below, where in the absence of regulation we argue that networks will often wish to set different charges for the two kinds of traffic. (However, in many cases, the welfare-maximizing charges will be the same for the two kinds of calls.) It is not clear whether this is an exogenous constraint faced by the networks or merely a strategic decision. (There was certainly no regulatory constraint that the two charges be equal.) An argument in favour of the former view is that arbitrage of some kind may force rates to be the same. For instance, if a network's MTM termination

market was effectively competitive—see Oftel, *Mobile Access and Call Origination Market: Identification and Analysis of Market and Determination of Market Power*, 3 October 2003.

[12] See Competition Commission (2003, Table 2.8). These figures are for a combined 900/1800 MHz spectrum network (i.e., the older networks operated by *Vodafone* and O_2), and include a "market share adjustment" which was not part of Oftel's own calculations. (The Commission otherwise broadly followed Oftel's model for calculating costs.) Oftel's corresponding figures were 5 pence per minute in 2002 and 3.4 in 2006.

charge is below its FTM charge, a fixed network might be able to "transit" its calls via another mobile operator and then pay only the MTM rate (plus a small transit charge). Alternatively, if MTM termination charges are set significantly below FTM charges, mobile subscribers might use their mobile phone to call other mobile subscribers instead of their fixed phone (when they are at home or in the office), even when the recipient subscribes to a rival network. (We previously discussed how on-net MTM calls can substitute for FTM calls.) In addition, a network whose FTM termination charge is regulated may be unwilling to choose a lower charge for MTM termination, if it suspects the regulator may use that information subsequently to tighten FTM regulation. On the other hand, it is plausible that ways round any technological problems could be found if the networks did wish to set differential charges. For instance, until 2004, mobile networks in New Zealand set differential rates (a positive charge for FTM termination and "bill-and-keep", i.e., a zero termination charge, for MTM traffic).[13]

Recall that the two older networks, O_2 and *Vodafone*, had their FTM termination charges explicitly regulated since 1998. Both networks were required to set their average weighted termination charges no higher than 11.7 pence per minute in 1999/2000, and then to implement real cuts in these charges of 9% per year for the coming two years. In particular, in comparison with the estimated cost of termination in 2000/01 of 7 pence per minute, the permitted termination charge was substantially in excess of the estimated cost. Thus, the issue for the 2002 enquiry was not just that the previous control regime was expiring, but also that it may have been too lax.

The newer networks, *Orange* and *T-Mobile*, were not explicitly regulated before the 2002 enquiry. However, they were informally controlled, in that Oftel indicated that if it were asked to make a determination on their FTM termination charges, it would set these to be roughly equal to those prevailing on the two regulated networks.[14] (The newer networks were allowed to charge slightly more since they used a different part of the spectrum with different cost characteristics.) As such, all networks set termination charges which were broadly similar, even though only two were formally regulated. In particular, the figures in Table 4 should not be regarded at an indication of what *unregulated* networks would charge.

2 A Basic Model of the Mobile Market

Here we present a basic formal model of the mobile telephony industry. The principal purpose of the model is to contrast the pricing of mobile-to-mobile (MTM) termination with the pricing of fixed-to-mobile (FTM) termination. We will compare what our model predicts

[13] In the enquiry, *Vodafone* suggested that regulation could apply differently to FTM and MTM termination, and there were no substantial technological obstacles to setting differential termination charges. See Competition Commission (2003, paras. 2.473–2.479).

[14] See Competition Commission (2003, para. 2.10).

to be the unregulated and the optimally regulated termination charges in each case.[15]

In this basic model we abstract away from call and network externalities, and we also suppose that mobile networks do not set different charges for on-net and off-net calls. (Section 3 will provide various extensions to this basic model.) In this case, provided FTM call charges are regulated to be equal to cost, we will see that welfare is maximized when both the FTM and MTM termination charges are equal to cost. Without regulation, we will see that mobile networks in this model will each wish to set an excessive FTM termination charge, while they are indifferent about their joint choice of MTM termination charge.

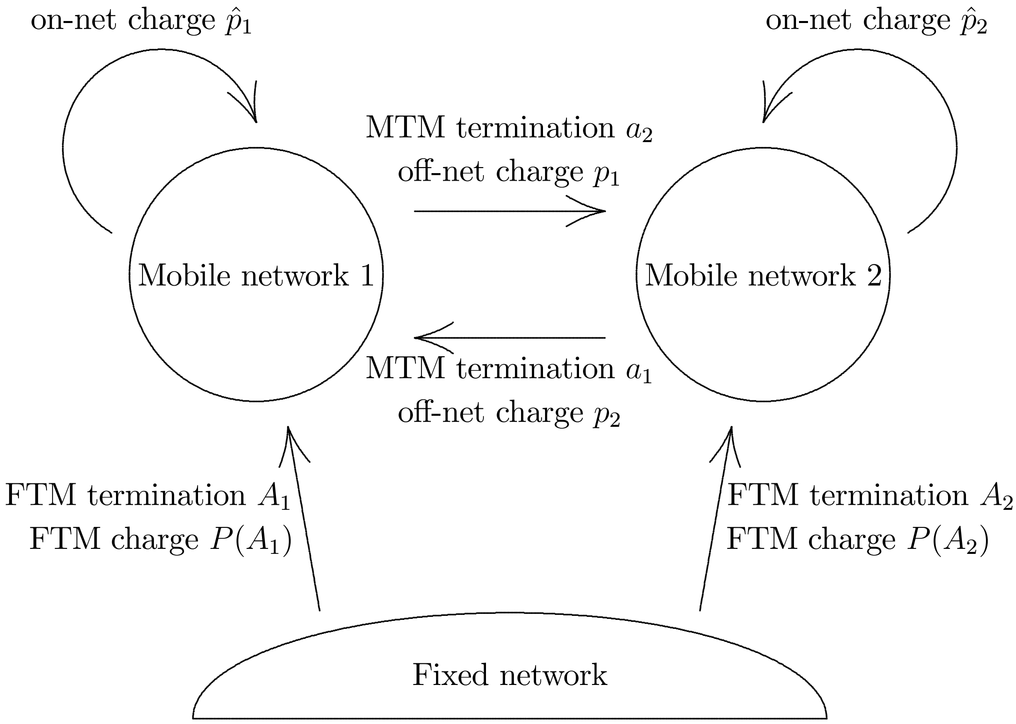

Figure 1: Call Termination on Mobile Networks

To model MTM calls, a standard model of two-way interconnection between symmetric networks is adopted.[16] Specifically, suppose there are two mobile networks, denoted $i = 1, 2$, which offer differentiated services. (In the next section, we allow for more than two mobile networks.) Each firm charges the other a termination charge for completing MTM calls originating on the other firm's network. The firms are assumed to negotiate a reciprocal

[15] The terms of reference for the 2002 enquiry explicitly required the Commission to consider termination charges in the absence of regulation—see Competition Commission (2003, Appendix 1.1).

[16] See Armstrong (1998), Laffont, Rey, and Tirole (1998a), and Carter and Wright (1999). In particular, we employ the model in section 8 of Laffont, Rey, and Tirole (1998a) where networks offer two-part tariffs to their subscribers.

MTM termination charge, denoted a. (The next section also allows for the possibility that each mobile network sets its MTM termination charge unilaterally.) For simplicity, all mobile subscribers are assumed to be identical in terms of their demand for calls to other mobile subscribers.[17] Specifically, if subscriber j faces a per-minute charge p for calling subscriber k, j will choose to make (an average of) $q(p)$ minutes of calls to k. Thus, each subscriber is equally likely to wish to call any other subscriber. For simplicity, suppose that mobile subscribers gain no benefit from receiving calls (from either the fixed or the mobile networks); the analysis is extend to allow for call externalities in section 3. Let $v(p)$ be consumer surplus associated with the demand function $q(\cdot)$, so that $v'(p) \equiv -q(p)$.[18]

Added to this framework is a fixed-line network, from which a demand for FTM calls is generated.[19] (As with the UK industry, we assume this fixed sector is separately owned from the mobile sector.) Each mobile network chooses a termination charge for completing FTM calls, and network i's FTM termination charge is denoted A_i.[20] We assume that in the first stage firms negotiate a reciprocal MTM interconnection price a, and subsequently, in stage two, they set their FTM termination charges A_i together with their retail tariffs to mobile customers.[21] If the retail price for FTM calls to mobile network i is P_i per minute, suppose that there are $Q(P_i)$ FTM minutes of calls to each subscriber on network i.[22] In general, we expect the price P_i to be an increasing function of the FTM termination charge A_i, and

[17] Clearly, this is a significant simplification. Indeed, the complexity of the tariffs used by mobile networks is intended in part to sort subscribers into groups with different calling patterns. The simplification is especially restrictive when we later look at market expansion: in the real world we expect that marginal subscribers are likely to have a reduced demand for calls compared to the average subscriber (although it is less clear that marginal subscribers will receive fewer calls than others). See Dessein (2003), Hahn (2004) and Valletti and Houpis (2005) for models where subscribers differ in their demand for calls. Dessein and Hahn show that the presence of heterogenous subscriber demand for calls does not make much difference to profit in this basic model where total market size is fixed.

[18] Throughout it is assumed the parameters are such that an equilibrium exists. Laffont, Rey, and Tirole (1998a) discuss the conditions for such an equilibrium to exist, which requires that the MTM termination charge does not differ too much from cost and/or that the mobile networks are sufficiently differentiated.

[19] This part of the model is mostly taken from section 3.1 of Armstrong (2002) and Wright (2002).

[20] Where possible, we use upper case notation for calls from the fixed network and lower case notation for calls from the mobile networks.

[21] The results obtained do not change when the FTM termination charge is instead set in stage two, with retail prices set in a subsequent third stage. However, the case with simultaneous setting of the FTM termination charge and other retail prices simplifies the exposition.

[22] Notice that we assume the fixed network can set different call charges to different mobile networks to reflect the networks' different termination charges. An alternative assumption is that the fixed network cannot 'price discriminate' in this way, perhaps because callers do not always know which mobile network they are calling. If so, the market failures identified in the following analysis are typically amplified. For instance, a small mobile network's termination charge has only an insignificant impact on the fixed network's average cost of providing FTM calls, and so the network has a unilateral incentive to set an extremely high termination charge, since such a charge does not cause demand for calls from the fixed network to its subscribers to fall significantly. Indeed FTM call charges may be *above* the monopoly level. See Gans and King (2000), Wright (2002) and Competition Commission (2003, paras. 2.48 and 2.136) for further details.

write $P_i = P(A_i)$. For instance, it may be that

$$P(A_i) = C + A_i \ , \tag{1}$$

where C is the fixed network's marginal cost of originating a call. Here, the FTM call charge is equal to the fixed network's total cost of making such calls. Such pricing could arise as a result of regulation of the fixed network or competition between fixed networks.[23] Let $V(P)$ be the consumer surplus function associated with the demand function $Q(\cdot)$, so that $V'(P) \equiv -Q(P)$.[24]

Each mobile firm is assumed to incur a marginal cost c_O of originating a call and a marginal cost c_T of terminating a call, so the actual marginal cost of a MTM call is $c_O + c_T$. In addition, there is a fixed cost f of serving each mobile subscriber, which includes the subscriber's handset, billing costs, and so on. For now, assume that FTM and MTM calls are independent markets, and that the call charge in one market does not affect the demand in the other market. (See section 3 for a discussion of how substitutability between the two types of calls may affect our analysis.) Figure 1 depicts our stylized model of the mobile industry.

In this basic model, we assume there is an exogenously fixed number of mobile subscribers, which we normalise to 1. (The next section allows for the possibility that the pool of subscribers will expand if the deals on offer improve.) The price of MTM calls from each network is assumed to be uniform, so that there is no difference between on-net and off-net call charges. Suppose that mobile network i sets the call charge p_i for its MTM calls, and charges a fixed (rental) charge r_i for subscribing. Then a subscriber's utility if she joins that network is

$$u_i = v(p_i) - r_i \ . \tag{2}$$

If we assume a Hotelling specification for subscriber choice, the market share of network i given the pair of utilities u_1 and u_2 is

$$s_i = \frac{1}{2} + \frac{u_i - u_j}{2t} \ , \tag{3}$$

[23] An earlier enquiry focussed on BT's FTM call charges, and recommended that BT's "retention"—its FTM call charge less the FTM termination charge payable—be directly regulated. See MMC (1999a) for further details. This regime expired in 2002, and after that BT's retention continued to be regulated, but now within its overall retail price cap. As such, Oftel said that "any reduction in termination charges by the [mobile networks] will be fully passed through one way or the other into retail prices to the [fixed network] customers, although such pass-through would not necessarily be into charges for fixed-to-mobile calls specifically" (Competition Commission, 2003, para. 2.42). However, in August 2006, BT's retail price controls were removed, and there is now no formal control of the firm's retention—see Ofcom, *Retail Price Controls*, 19 July 2006.

[24] One might also consider mobile-to-fixed calls. However, in the simple frameworks we present below, these calls play no significant role in the analysis, and are ignored from now on. (A mobile network would just set the price for such calls equal to its cost of providing such calls in the models we present.) Similarly, fixed-to-fixed calls play no role in the analysis.

where t is the 'transport cost' parameter which represents the competitiveness of the retail market for mobile subscribers.

Define

$$F(A) \equiv (A - c_T)Q(P(A)) \tag{4}$$

to be a mobile network's profit, per subscriber, from providing termination services for the fixed network when its FTM termination charge is A. Network i's profit is then

$$\pi_i = s_i \times \left[\underbrace{r_i - f + (p_i - c_O - s_i c_T - s_j a)\, q(p_i)}_{\text{retail services}} + \underbrace{s_j\,(a - c_T)\, q(p_j)}_{\text{MTM termination}} + \underbrace{F(A_i)}_{\text{FTM termination}} \right] . \tag{5}$$

This profit consists of the retail profit from supplying service to its subscribers, the profit from providing termination for the rival mobile network, and the profit from providing termination for the fixed network.

2.1 Fixed-to-mobile call termination

From expression (5), each mobile network will sets its FTM termination charge A_i to maximize its profits from FTM call termination, $F(\cdot)$. This is a dominant strategy for each network, regardless of choices for retail tariffs and the MTM termination charge. By setting A_i to maximize F, each firm will be able to subsidize subscribers to the maximum extent, thereby increasing market share without needing to lower profit per subscriber. We denote the resulting unregulated termination charge by A_M, to indicate it is the termination charge that would be chosen by a monopoly mobile network. This situation, in which the equilibrium termination charge with competition is equal to the monopoly charge, is sometimes known as a *competitive bottleneck*.[25] The result does not depend on the competitiveness of the market for subscribers (as measured by t). Thus it is perfectly possible that one part of the mobile market (the market for mobile subscribers) is highly competitive, yet the other part (FTM call termination) is essentially a pair of monopolies.

In the 2002 enquiry, Oftel stated that "it was quite possible that the profit-maximizing [FTM termination charge] would be substantially above the current level, possibly exceeding 20 ppm." *Vodafone* stated that the unregulated level was in range 17 ppm to 20 ppm. At the time these estimates were given, the cap on O_2 and *Vodafone*'s termination charge was 9.3 ppm, so *Vodafone* was suggesting that the FTM termination charge would roughly double without the charge control.[26]

[25] See Armstrong (2002) and Wright (2002) for more detailed analysis. Armstrong (2006) shows how a similar analysis can be applied to other industries, including advertising in the media, shopping malls and supermarkets. The key feature in these markets (as modelled) is that one side of the market chooses to interact with just one intermediary, while the other side interacts with several intermediaries in order to gain access to a wide selection of agents on the other side. Thus, an intermediary hold a monopoly over providing access to those agents who deal with it exclusively, and accordingly sets a high price to the other side.

[26] See Competition Commission (2003, paras. 2.440–2.445). Somewhat implausibly, the other three mobile networks claimed that they would not raise their termination charges if the charge control was lifted.

Suppose social welfare is measured as the simple unweighted sum of consumer surplus and profit.[27] When there are no call or network externalities, welfare (as measured by the sum of consumer surplus and profit) in the FTM segment when the termination charge is A is

$$\underbrace{V(P(A))}_{\text{consumer surplus}} + \underbrace{F(A)}_{\text{mobile termination profit}} + \underbrace{[P(A) - C - A]\, Q(P(A))}_{\text{retail profit by fixed network}} , \tag{6}$$

which simplifies to

$$V(P(A)) + [P(A) - (C + c_T)]\, Q(P(A)) \ .$$

As one would expect, this is maximized by setting the FTM call charge equal to the cost of such calls:

$$P(A_W) = C + c_T \ ,$$

where A_W denotes the welfare-maximising FTM termination charge. This ensures fixed-line callers face a FTM price equal to the true marginal cost of their calls, namely $C + c_T$. In the particular case in which the FTM call charge is equal to the fixed network's call cost, so that (1) holds, welfare is maximized with a FTM termination charge equal to cost:

$$A_W = c_T \ .$$

On the other hand, if $P(A) > C + c_T$ then welfare would be maximized by setting $A < c_T$ to counteract the markup present in the FTM call charge.

2.2 Mobile-to-mobile call termination

Suppose that the mobile networks each set the FTM termination charge A. (As discussed, without regulation we expect to see $A = A_M$. However, regulation may impose $A = A_W$, or some other charge, in which case it is still useful to analyze the remaining parts of the mobile market.)

Suppose for now that the two mobile networks coordinate on a choice of MTM termination charge, say a. At a symmetric equilibrium, both mobile networks will set their MTM call charge equal to $p(a)$, say, when the MTM termination charge is equal to a. Here, the call charge is equal to the average cost of making calls:[28]

$$p(a) = c_O + \tfrac{1}{2}(c_T + a) \ . \tag{7}$$

[27] However, the mobile operators appeared to view the objective of policy to be to maximize mobile subscriber numbers or the number of calls involving mobiles—see Competition Commission (2003, para. 2.404), for instance.

[28] See Proposition 7(ii) in Laffont, Rey, and Tirole (1998a) for details. This is at odds with a strange statement made by the Commission (Competition Commission, 2003, para. 2.50): "changes in mobile-to-mobile termination charges do not affect the aggregate finances of the industry as a whole, and for this reason may prima facie be expected to have no effect on [the mobile networks'] retail prices."

Similarly to (4), write

$$S(a) \equiv (p(a) - c_O - c_T)q(p(a)) \tag{8}$$

for the mobile sector's per-subscriber profit from supplying MTM calls, where $p(a)$ is given in (7). Expression (5) then simplifies to

$$\begin{aligned} \pi_i &= s_i \times [r_i - f + S(a) + F(A)] \\ &= \left[\frac{1}{2} - \frac{r_i - r_j}{2t}\right] \times [r_i - f + S(a) + F(A)] \ , \end{aligned} \tag{9}$$

where we used the market share expression (3). Network i will then choose r_i to maximize (9). At a symmetric equilibrium, each firm will set the rental charge r, where r satisfies

$$r - f + S(a) + F(A) = t \ . \tag{10}$$

From (9) and (10), it follows that in equilibrium the mobile industry obtains profit t, regardless of the two termination charges a and A. Thus, in this very simple setting firms are indifferent to the choice of MTM termination charge. As such, it is plausible they will not be strongly opposed to regulatory intervention to set this termination charge at the socially optimal level. Since the socially optimal MTM call charge is equal to the cost $c_O + c_T$, expression (7) implies that the efficient MTM termination charge is $a = c_T$, just as with FTM termination. In particular, optimal policy treats the two termination charges symmetrically.

2.3 Discussion

The result expressed in (10) is striking. For any MTM termination charge a and any FTM termination charge A, competition between networks to attract subscribers will ensure industry profit is always equal to the competitiveness parameter, t. If firms can obtain high profits from terminating calls from the fixed network, or by setting high MTM call charges, this makes the firms compete harder for subscribers—each subscriber brings a profit $S(a)+F(A)$—and this acts to pass these profits onto the subscribers themselves. Correspondingly, if regulation squeezes out excess profits in one segment (such as FTM call termination), this will lead to price rises for mobile subscribers.[29] (Specifically, expression (10) suggests that the rental charge will rise by a factor equal to the reduction in per-subscriber FTM termination profits.) This result is sometimes termed the "waterbed" effect: if one source

[29] The Competition Commission (2003, para. 1.13(b)) did not expect retail prices to rise if call termination charges were reduced by regulation: "The [mobile networks] need not increase their retail prices to restore revenue lost through the capping of termination charges, as their business plans project a continued downward trend in retail prices and they could recoup these revenues by reducing the rate of retail price reductions for a period." Nevertheless, the main point, that reducing the profits from call termination is likely to lead to higher retail prices than would otherwise be the case, remains.

of a network's per-subscriber profit is reduced (e.g., by regulation), that will cause the networks to compete less hard for subscribers, and the overall impact of the profit reduction is mitigated or eliminated altogether.[30]

Expression (10) suggests some symmetry in the mobile sector's incentives to set the FTM and the MTM termination charges. This is only partly true. The two charges share the feature that firms are indifferent to each of their levels, if all firms choose the same level. In particular, in this basic model firms should not object to regulatory intervention to bring *each* firm's FTM termination charge down from the unregulated level A_M to the socially optimal level A_W. However, the central difference between the two kinds of termination charge is that, without regulation, in this model MTM charges are chosen by negotiation between mobile firms, whereas, without regulation, FTM charges are chosen unilaterally by mobile firms. (BT's mobile-to-fixed termination charge is fixed by regulation, and so there is no 'negotiation' between a mobile firm and the main fixed operator. In section 3.5 we discuss the hypothetical case where fixed and mobile networks do negotiate over their termination charges.) If mobile firms instead set their MTM termination charges unilaterally rather than by mutual negotiation, those charges would also be chosen to be excessively high to exploit each firm's competitive bottleneck (see section 3.2 below). But there is no natural way to make mobile firms choose their FTM termination charge as a group (in which case they would be indifferent to the choice). If firms choose this charge unilaterally, it is a dominant strategy for them to set it equal to the monopoly level A_M. Indeed, a firm would be at a competitive disadvantage if it chose a lower charge while its rival set the monopoly charge, since its rivals would then have greater ability to fund subsidies to attract subscribers.

Nevertheless, the four mobile networks *did*, in broad terms, object collectively to such a reduction in the FTM termination charge, which suggests that this basic model with a 100% waterbed effect fails to capture important aspects of the real-world market. See section 3.3 below for a model in which mobile firms do collectively prefer a higher to a lower FTM termination charge. In addition, even if networks make the same equilibrium profits with different choices of termination charge, this does not imply that a *transition* from one equilibrium to another (e.g., as mandated by new policy) will not be costly to firms. For instance, if networks had long-term contracts with subscribers, designed at a time when termination revenues were relatively high, then if policy reduces termination revenues, firms may make losses on these subscribers until contracts expire.[31]

A feature of the waterbed effect is that unregulated monopoly profits from FTM call

[30] In this basic model, there is a '100%' waterbed effect, in that reduced profits from one source are completely clawed back from subscribers so that the overall profit impact is zero. Alternative ways of modelling competition for subscribers might lead to less than 100% (or perhaps greater than 100%) waterbed effects. See for instance the discussion in Competition Commission (2003, pages 292-294), as well as the model in section 3.3 below. In a panel data study, Genakos and Valletti (2007) uncover a strong long-run waterbed effect resulting from regulated changes in termination charges. More generally, the extent of passthrough of termination revenue can be derived in alternative oligopoly settings by using the techniques in Dixit (1986), for instance.

[31] See Competition Commission (2003, para. 2.530).

termination are passed onto mobile subscribers in the form of low rental charges (e.g., "free" handsets and the like). To the extent this occurs, the market failure associated with FTM termination does not lead to excessive profits by mobile networks, but rather to a sub-optimal pattern of prices. Of course, this observation should not affect the welfare analysis: if high margins on FTM call termination lead to negative margins on services to mobile subscribers, there is allocative inefficiency regardless of whether overall profits in the mobile sector are excessive or not.[32] Relatedly, the mobile firms "advanced the argument that, because most people had a mobile phone, what they lost in high termination charges they gained in low access and outbound call charges".[33] However, even if all fixed-line subscribers also had a mobile phone, this argument is not correct: since high termination charges lead to allocative inefficiency, the total "size of the cake" is shrunk, and the gain from handset subsidies and the like is *smaller* than the losses caused by high FTM call charges.[34] Moreover, in this particular model, the competitiveness of the retail mobile market has little bearing on the welfare costs of unregulated FTM termination charges. (In our basic model, this competitiveness is represented by the differentiation term t, and this parameter does not enter into the expression for welfare in (6).)

A somewhat related point is made by several of the mobile networks, which is that their tariffs would approximate "Ramsey" prices in the absence of regulation.[35] The argument is that networks offer subscribers a bundle of services, including inbound and outbound calls, in return for a specified tariff. Competition between networks forces them to offer subscribers the best deal consistent with the networks earning a reasonable profit. (Prices which maximize aggregate consumer surplus subject to the firm earning a reasonable profit are termed Ramsey prices.) One half of this argument is perfectly valid: if the retail market is reasonably competitive then we expect firms to compete by offering a tariff package which tends to maximize their subscribers' surplus subject to achieving reasonable profit.[36] However, the argument completely ignores the fact that mobile subscribers do not pay for their incoming calls (or choose the volume of such calls); that is to say, competition may well ensure that networks maximize their subscribers' surplus, but it ignores the consumer

[32] Competition Commission (2003, para. 1.8(d)) noted that "the excess charges for termination have the further effect that they serve to encourage or facilitate significant distortions in competition because [mobile networks] are not obliged to charge and subscribers are not obliged to pay the economic cost of handsets. This lead to the undervaluation of mobile phone handsets by the [...] customers combined with a greater turnover ('churn') that would take place if customers paid charges which reflected the proper valuation of such handsets." In section 3.4, where we discuss substitution between MTM and FTM calls, we will see there is a potential distortion in competition between FTM and on-net MTM calls when FTM termination charges are set at too high a level.

[33] See Competition Commission (2003, paras. 2.390 to 2.400).

[34] The Competition Commission (2003, paras. 2.550 to 2.558) suggest that the impact on overall welfare (that is, consumer surplus plus profit) of their proposed reduction in termination charges is of the order of £325m to £700m over the period of the proposed regulation.

[35] See Competition Commission (2003, paras. 2.429 to 2.446).

[36] See section 3 of Armstrong and Vickers (2001) for a formal model with this feature.

surplus of callers from the fixed network to mobile subscribers. Indeed, without regulation, competition *forces* the mobile networks to exploit fixed callers in order to be able to attract mobile subscribers. There is no reason at all to expect that, in the absence of regulation, the structure of outbound and inbound call charges will approximate Ramsey prices.[37]

A final point about the waterbed effect is that it suggests that there will be a positive relationship between the surplus enjoyed by mobile subscribers and the level of FTM termination charges. In particular, evidence that mobile subscribers are satisfied with the deals they receive (or that mobiles prices are unusually low compared with those in other countries) should *not* necessarily be taken as evidence that the overall telecommunications market is performing well.

There is no reason why only "large" (or incumbent) mobile networks should have their FTM termination charges regulated, while "small" (or new) networks are free to set whatever charge they wish. A small network will set a monopoly FTM termination charge in exactly the same way as a large firm. Indeed, if incumbent firms are regulated while new firms are not, new firms have a powerful advantage over incumbent firms, in that they have greater ability to fund subsidies to subscribers. This asymmetric treatment of firms in the industry is likely to lead to a distorted pattern of supply. A recent illustration of what happens if some firms are regulated and some are not is presented in the following table, where the figures (in pence per minute) state the (now) five mobile networks' termination charges in March 2006 to be:[38]

Table 5: The Impact of Regulation on Termination Charges

	Daytime	Evening	Weekends
O_2	6.4	6.3	3.1
Orange	7.6	5.4	4.3
T-Mobile	8.1	4.0	4.0
Vodafone	8.5	3.4	2.8
H3G	**15.6**	**10.8**	**2.5**

At around the time of the 2002 enquiry there was a fifth entrant, *H3G*, whose termination charges in 2006 were not directly controlled. As is clear, on weekdays at least, this firm took advantage of its unregulated position to set charges which are roughly double those levied by its regulated rivals.

An implicit assumption in the basic model is that fixed-line callers alone determine the number and length of FTM calls. If in practice the number and length of FTM calls is jointly determined by the caller and receiver, then high FTM termination charges could lead mobile networks to *pay* subscribers for receiving FTM calls (rather than merely reducing rental

[37] This basic market failure was understood and acknowledged by at least one of the mobile networks, *Vodafone*. (See Competition Commission, 2003, paras. 2.80 and 2.107.)

[38] Taken from Figure 1 in Ofcom (2006b).

charges), so as to promote more FTM calls. For instance, in January 2006 *H3G* announced that it would pay its subscribers 5 pence per minute for receiving calls, a marketing tactic which surely is motivated by its highly profitable termination charges (see Table 5).[39]

This point is also helpful in understanding the difference between the "caller-pays" regime as used in the UK, in which FTM termination charges and call charges are relatively high, and a "receiver-pays" regime as used in the United States, whereby the price of a FTM call is low but mobile subscribers incur charges for receiving calls. In the United States, reciprocity requirements mean that FTM termination charges are very low, quite likely lower than the cost of termination. With little termination revenue, mobile operators may choose to charge mobile subscribers directly to recover their costs and/or to induce their subscribers to discourage incoming traffic. For instance, with a charge for receiving calls, a subscriber may keep his or her phone switched off to eliminate incoming calls, something which is often socially inefficient. The key point is that it is not that a receiver-pays regime helps to solve the problem of high FTM termination charges, but rather that a regulatory requirement to set unduly low FTM termination charges may induce networks to charge their subscribers for receiving calls.

3 Extensions to the Basic Model

3.1 Call externalities and price discrimination

In this section, we extend the benchmark model in two ways. First, we allow mobile subscribers to obtain benefits from receiving calls. Second, we allow mobile networks to set differential off-net and on-net MTM call charges.[40] To simplify the analysis of MTM calls, we continue to assume an exogenously fixed number of mobile subscribers.

As depicted in Figure 1, denote firm i's on-net MTM price by $\hat{p}_i$ and its off-net MTM price by p_i.[41] If firm i's market share is s_i, its subscribers make a fraction s_i of their calls on-net and the remaining $1-s_i$ calls off-net. Suppose each subscriber obtains a linear benefit from receiving calls from other mobile subscribers and a (possibly different) linear benefit from receiving calls from the fixed network. Specifically, a mobile subscriber obtains a benefit

[39] However, there are limits to the use of reception payments. In Italy, when mobile operators tried paying subscribers for calls received, they found people were calling their own mobile phones from office lines to obtain these rebates.

[40] The setting used is similar to Laffont, Rey, and Tirole (1998b), Gans and King (2001) and Berger (2005), extended to take into account FTM calls and the presence of more than two mobile firms. Notice that the presence of call externalities *without* price discrimination between off-net and on-net calls does not add much of interest to the benchmark analysis. For instance, see Armstrong (2002, pages 366-367).

[41] In the following analysis, we are probably over-optimistic about the extent to which subscribers are aware of price differences in off-net and on-net calls. For instance, see Competition Commission (2003, para. 2.136) for evidence that only a small fraction of subscribers were likely to know whether they were calling a mobile phone on the same network as themselves or not. On the other hand, Table 3 above suggests that consumers do respond to price differentials.

$BQ + bq$ if she receives Q calls from the fixed network and q calls from the mobile networks. Here B and b measure the strength of the respective call externalities.[42]

In the basic model of section 2 we assumed a duopoly in the mobile market, since no extra insights are obtained by allowing for more than two firms. When discussing the effect of price discrimination, however, it turns out that additional understanding is gained by allowing for an arbitrary number of mobile networks. To that end, suppose there are $K \geq 2$ symmetrically placed mobile networks.

The details of this analysis are technical, and we present them in Appendix A. In summary, though, the impact of call externalities on the choice of the FTM termination charge is relatively straightforward. Without call externalities, unregulated firms would like to choose A to maximize $F(A)$, but with the externality B, firms would like to choose to A to maximize

$$BQ(P(A)) + F(A) \ . \tag{11}$$

Of course, the unregulated FTM termination charge will be lower than without call externalities. If a network sets a high FTM termination charge, this will reduce the volume of calls received by its subscribers from the fixed network, and hence reduce the rental fee it can charge its subscribers if it wishes to maintain market share. In effect, the call externality B has exactly the same effect on a network's choice of FTM termination charge as a reduction in its termination cost c_T by B. Thus, as emphasized by the networks in the 2002 enquiry, call externalities do mitigate a network's incentive to set high FTM termination charges.[43]

Welfare in the FTM segment when the FTM termination charge is A is equal to

$$\underbrace{V(P(A))}_{\text{consumer surplus}} + \underbrace{BQ(P(A))}_{\text{call externality}} + \underbrace{F(A)}_{\text{mobile termination profit}} + \underbrace{[P(A) - C - A]\, Q(P(A))}_{\text{retail profit by fixed network}} ,$$

which is maximised by setting the FTM call charge equal to the 'adjusted' cost of FTM calls:

$$P(\hat{A}_W) = C + c_T - B \ . \tag{12}$$

This ensures fixed-line callers face a FTM price equal to the true marginal cost of their calls, adjusted to take account of the impact of their calls on mobile recipients. When (1) holds,

[42] The enquiry barely discussed call externalities at all—see Competition Commission (2003, paras. 8.257 to 8.260). But it is beyond doubt that call externalities are significant, since why else would anyone leave their mobile phone on to receive calls? But the claim made (see para. 8.257) is that call externalities are "largely internalized as people tended to be in stable calling relationships with each other. The caller might be prepared to make a call even if his expected benefit was less than the price, because he expected that further call (or calls) would be generated, initiated and paid for by the other party, from which he would receive a benefit without having to pay." This argument suggests that welfare-optimal call charges need not be significantly subsidised below cost since these regular calling pairs will tend to make each call when the *sum* of the two parties' call benefits are greater than the charge for the call. This may well be true. However, this argument has little relevance for the model presented in this section, where networks set relatively high off-net prices in order to reduce the number of calls that subscribers on rival networks receive. Indeed, call externalities are probably one of the reasons for the major off-net/on-net call charge disparities we observe.

[43] For instance, see Competition Commission (2003, para. 2.109).

welfare is maximized with a FTM termination charge equal to the adjusted cost

$$\hat{A}_W = c_T - B \ . \tag{13}$$

In sum, when mobile subscribers obtain benefits from receiving calls from the fixed network, this causes *both* the unregulated equilibrium termination charge and the welfare-maximizing termination charge to fall. As such, the presence of such externalities does not diminish the potential need to regulate these termination charges.

Turning next to the MTM termination charge, in the specific model one can show that when this charge is set equal to a, each network sets its on-net price $\hat{p}$ and off-net price p in the following manner:

$$\hat{p} = c_O + c_T - b \ ; \ p(a) = c_O + a + \frac{1}{K-1} b \ . \tag{14}$$

Thus, the on-net call charge is equal to a network's cost for the call, $c_O + c_T$, adjusted downwards to reflect the call externality its subscribers enjoy from being called more often from other people on the same network. This on-net call charge is not affected by a, and is equal to the socially efficient call charge. The off-net call charge $p(a)$ is equal to a network's cost for the call, $c_O + a$, adjusted *upwards* to reflect the fact that when its subscribers make fewer calls to subscribers on rival networks, call externalities imply that its rivals' abilities to compete are harmed and this benefits the original network. This represents the chief anti-competitive motive to set high off-net call charges.

Notice that, even if the MTM termination charge is set equal to cost ($a = c_T$), the prices in (14) differ for on-net and off-net calls. Thus, this model predicts that observed differences in these call charges are by no means solely due to above-cost MTM termination charges. Nevertheless, it is not easy to reconcile the predicted prices in (14) with the actual, rather extreme, differences described in Table 2.[44] In the enquiry, the mobile networks themselves did not provide compelling reasons for why they set such different prices,[45] and so at present it remains a puzzle why this pattern of prices was—and to a large extent, still is—observed. In section 3.2, we will show that when networks set their MTM termination charges unilaterally, this will lead to higher termination charges and this in turn will feed through into higher off-net call charges. However, this does not help explain the high differentials in Table 2, which took place when the MTM termination charges were regulated.

As in section 2, the waterbed effect implies that profit from FTM termination has no impact on the equilibrium profit. However, unlike the previous analysis, here mobile networks

[44] Even if termination charges exceed cost c_T by 5 pence per minute, which in section 1 we show is roughly what the Competition Commission concluded at the time of the enquiry, to obtain a difference between p and $\hat{p}$ of nearly 20 pence per minute would with four firms require a call externality of an implausible size (around 11 pence per minute), a size which would probably make $\hat{p}$ in (14) negative.

[45] See Competition Commission (2003, paras. 2.124-2.125).

are not indifferent to the joint choice of MTM termination charge.[46] In the model, one can show that the mobile industry's profit is maximized when the MTM termination charge is chosen so that the off-net call charge is

$$p = c_O + c_T - \frac{K}{K-1}b + \frac{1}{K-1}\frac{q(p)}{q'(p)} \ . \tag{15}$$

Notice this call charge is *lower* than the welfare-maximizing off-net call charge, which (like the on-net call charge) is equal to $c_O + c_T - b$. Thus, in contrast to incentives concerning the FTM termination charge, here the danger is that mobile networks will coordinate on a MTM termination charge which is too low. In particular, in this framework the off-net call charge p is predicted to be set below the on-net charge ($\hat{p} = c_0 + c_T - b$), which is at odds with real-world observation (as in Table 2 above).

Why do we predict, in this model, that firms will agree to set a particularly low MTM termination charge if free to do so? Unless firms set a low termination charge, the call charges in (14) will be such that it is more expensive to call off-net than on-net. In such a situation, subscribers will, all else equal, prefer to join a larger network since they can then make a larger fraction of their calls at the cheaper rate. In other words, the market will exhibit (positive) network effects. As is well known, in such markets competition is particularly fierce and profits are low. In principle, firms can overturn this effect by setting a low termination charge, which implements off-net call charges which are below on-net charges. In this case, subscribers will, all else equal, prefer to join a smaller network, which acts to relax competition. Thus, in principle, there is a need for regulation to achieve an efficient MTM termination charge. Unlike the case of FTM termination, though, regulation should here ensure that the MTM termination charge not be set at too *low* a level.

However, and in contrast to the FTM case, this potential market failure is reduced when there are more mobile firms in the market. Expression (15) shows that when K is reasonably large, networks will voluntarily jointly choose a termination charge which approximates the socially optimal termination charge. We deduce that when there is a reasonable number of mobile firms in the market, the private incentives to choose the MTM termination charge are approximately in line with overall welfare. As such, the modest benefits associated with regulatory control of the MTM termination charge may be outweighed by the various costs associated with regulatory intervention. Moreover, if subscribers gain approximately the same benefit from being called from the fixed network as from a rival mobile network (i.e., $B \approx b$), then when K is large the two socially optimal termination charges are approximately equal. That is to say, it makes little difference to optimal policy if networks are forced (e.g., by technology or arbitrage constraints) to set a uniform charge to terminate all traffic.

[46] This is at odds with some statements made by the Commission in their report. For instance, Competition Commission (2003, para. 2.51) states that if call traffic between mobile networks is balanced (as it is in our formal model), "the finances of the individual [mobile networks] would be unaffected whatever the level of mobile-to-mobile termination charges, except perhaps at extremes far beyond any level discussed in this report."

To the best of our knowledge, no regulator has taken seriously the concern that firms have an incentive to set MTM termination charges which are too low. If anything, the concern has been the opposite. One reason may be that technology or arbitrage force firms to set the same charge for terminating both FTM and MTM traffic. Another reason why the four existing operators may have jointly preferred high MTM termination charges is that, at the time of the enquiry, a fifth network had just entered the market. By setting above-cost MTM termination charges, the incumbent networks can induce firm-specific network effects which make entry less attractive for the newcomer. With high MTM termination charges, off-net calls will be more expensive, which particularly hurts a small network since the bulk of its customers' calls will be off-net. Call externalities reinforce this effect, since when the established firms have high off-net prices, subscribers of a new (smaller) network will also receive relatively few calls.[47]

The next section provides an additional reason why MTM termination charges might be set at too high a level in the absence of regulation.[48]

3.2 Unilateral choice of MTM termination charges

Up until this point we have assumed that mobile networks negotiate a reciprocal MTM termination charge. Because networks were modelled as being symmetric, this negotiation process was trivial, and firms were assumed to choose a to maximize their joint profit (taking as given the downstream non-cooperative rivalry to attract mobile subscribers). However, it is by no means clear that unregulated networks do actually negotiate over their mutual MTM termination charges. For instance, in the 2002 enquiry *Vodafone* suggested that MTM termination should not be explicitly regulated, but rather that firms should be required to enter into bilateral negotiations over reciprocal termination charges, a suggestion which was resisted by the other mobile networks.[49] Alternatively, if each network must set the same charge for terminating traffic from all sources (due to technological or arbitrage constraints), then the incentive to exploit callers from the fixed network may over-ride subtle considerations about relaxing competition for mobile subscribers by means of bilateral negotiations with other mobile networks. (Recall from section 1 that the majority of terminating traffic originated on the fixed network.) In addition, asymmetries between mobile networks may make it harder for firms to reach a negotiated agreement. In sum, the relevant outcome without regulation might not be the coordinated decision for MTM termination charges (as

[47] Hoernig (2007) analyses the impact of on-net and off-net price differentials on the profitability of small networks. He shows that larger firms will choose greater differentials than smaller firms.

[48] At lest two further explanations have been proposed in the literature. First, Cherdron (2002) considers a setting where calling patterns are biased towards peers (closed user groups). Setting above cost MTM termination charges is a way to endogenously differentiate the networks, so that consumers prefer to stick to the network that their peers join, thereby reducing competition between operators. Second, Höffler (2006) argues in a dynamic model that higher MTM termination charges can sustain tacit collusion between firms for a wider range of discount rates.

[49] See Competition Commission (2003, para. 2.473).

we have assumed until now), but rather the outcome corresponding to when each network chooses its MTM termination charge unilaterally.

The analysis of such a model is complex, and we provide details of the duopoly case in Appendix B.[50] Broadly speaking, when network i raises its MTM termination charge there are four effects: (i) its profit from supplying call termination to its mobile rivals increases (just as with FTM termination); (ii) its rival is forced to raise its own off-net call charge, which places it at a competitive disadvantage compared to firm i; (iii) firm i's subscribers will receive fewer calls from the rival's subscribers due to (ii) (just as in the FTM case), which harms i's ability to compete for subscribers when call externalities are important, and (iv) it amplifies the differential pricing between on-net and off-net calls in the market which, as we argued in section 3.1, tends to intensify competition for subscribers. The equilibrium choice of termination charge will trade off these four effects. Effects (i) and (ii) together suggest that a firm will want to set its termination charge *above* the monopoly level, since (ii) gives a reason to boost this charge in addition to extracting termination profits. But (iii) and (iv) put some downward pressure on termination charges. Note that our analysis of FTM termination charges needs to consider just effects (i) and (iii).

In the model in the appendix, we ignore call externalities (so that (iii) plays no role) and show that the net incentive is to set a termination charge above the efficient level but below the monopoly level. Thus, in contrast to the case where the termination charge was jointly chosen by the mobile networks (where the danger was that firms would choose too low a charge), here the (more intuitive) danger is that the charge will be too high. In an extension to the model in the appendix, we allow for more than two mobile firms, and find that when the number of firms is large, each firm will set its unilateral MTM termination charge at approximately the monopoly level.[51] This suggests that effects (ii) and (iv) vanish when the market becomes more fragmented, and we are left only with (i). This contrasts with the analysis presented in section 3.1, where we argued that with many mobile networks in the market, firms jointly have an incentive to choose approximately the efficient MTM termination charge.

3.3 Network externalities and market expansion

A potentially important feature of the mobile market is that the number of mobile subscribers is not constant. Indeed, section 1 reports that the market grew dramatically in the years before the 2002 enquiry, although it is plausible that the market now is close to saturated. In this section, we investigate how the analysis presented in section 2 is modified when market size is affected by the utility subscribers receive.

Again, the technical details of the (complicated) analysis are relegated to Appendix C,

[50] The analysis is closely related to section 3 in Gans and King (2001) and Behringer (2006). Related analysis is found in section 6 of Laffont, Rey, and Tirole (1998a) for the case of linear tariffs and no on-net and off-net price differentials.

[51] Details are available from the authors on request.

where we present a duopoly model without call externalities and without price differentials between on-net and off-net calls.[52] (The presence of call externalities will amplify the impact of network externalities, since users will receive more calls when there are more mobile subscribers.) The first point is that the possibility of market expansion does not mitigate a mobile network's incentive to choose a high FTM termination charge. Their incentive is still to extract as much profit from the fixed callers as possible (adjusted to take account of call externalities if relevant). Thus, the unregulated FTM termination charge does not depend on the whether or not the market is saturated. However, we show that market expansion possibilities do suggest that the efficient FTM termination charge is above cost (unlike in the basic model). The reason is that high FTM termination profits ($F(A)$) will feed through into low tariffs for mobile subscribers via the waterbed effect, which will in turn feed through into more people deciding to subscribe to mobile networks, which will in turn benefit all users (fixed and mobile) since they have more people to call.

Nevertheless, the efficient FTM charge still lies below the unregulated level. The unregulated FTM charge is the charge which maximizes a mobile firm's profit from providing call termination to the fixed network. Starting from this profit-maximizing charge, a small reduction in the charge has a second-order impact on the profit from call termination, and hence only a second-order impact on the number of mobile subscribers. However, it has a first-order impact on the price for FTM calls. Therefore, welfare certainly rises with a small reduction in the FTM charge from the unregulated level. In sum, there remains a (reduced) rationale for regulatory control of the FTM termination charge.

The Commission estimated that the "externality surcharge" needed to reflect the benefits of market expansion was 0.45 pence per minute, or around 8% of their estimated costs of call termination (in 2003).[53] (The Commission did not distinguish between FTM and MTM termination charges in their analysis of this issue.) Their chosen methodology was to assume that the mobile networks could, at least in part, target a subsidy to new subscribers rather than to the body of existing subscribers.[54] The Commission then estimated the subsidy which, when so targeted, "brings about at least as much external benefit as the amount of

[52] The mobile-to-mobile part of the following discussion is closely related to section 5 of Dessein (2003). (He focusses his analysis on small deviations in the MTM termination charge from cost, which is related to our focus on the "almost saturated" market case.) Like us, he shows that the unregulated MTM charge is below cost. The finding derived below, that the welfare-maximizing FTM termination charge is above cost but below the unregulated level, has been noted in the existing literature (Armstrong (2002), Wright (2002), and Valletti and Houpis (2005)), but previous models have unrealistically assumed away MTM calls. Thus, the model we present here allows us to take a first step towards understanding whether it is FTM or MTM termination charges, or both, which should contribute to the subsidization of mobile subscription. In future work, it would be valuable to extend the analysis we present here to the more realistic case with on-net and off-net pricing differentials.

[53] See Competition Commission (2003, paras. 2.334-2.386). This surcharge is similar in size to that imposed after the earlier 1998 enquiry (see Competition Commission, 2003, para. 2.336).

[54] However, it is not clear how exactly this subsidy could be so targeted—see Competition Commission (2003, para. 2.360). In the model in the appendix, we assume that the subsidy is applied equally to all mobile subscribers.

the subsidy" (para. 2.373). Needless to say, such a calculation is fraught with difficulties to do with the reliability of estimates of the various elasticities and externalities, and the mobile networks may have an incentive to exaggerate the importance of external benefits.

What about the choice of MTM charge? In the appendix, we show that firms would like to choose a termination charge which is below cost, just as in the situation with call externalities and price discrimination in section 3.1. The reasons why firms wish to set $a < c_T$ in the two frameworks are subtly distinct. In the previous analysis of call externalities and price discrimination, a low MTM termination charge served to transform the market for subscribers into a market with negative, firm-specific network effects. Because the off-net call charge was then lower than the on-net call charge, subscribers preferred to join the smaller network. This acted to relax competition for subscribers, since a price-cut by one network did not bring forth a large gain in market share. In the current context, with the potential for market expansion and no price discrimination, there is an industry-wide network effect: a price-cut by one firm expands the overall market, and this benefits the rival's subscribers too. Such markets are also not too competitive, since a price-cut by one firm again does not bring forth a substantial gain in market share. Moreover, this industry-wide network effect is amplified if the MTM call charge is low, since an additional subscriber causes a greater external benefit to each existing subscriber. (This is seen formally in expression (28).) Therefore, competition is further relaxed if firms offer a low MTM call charge, i.e., if they agree to set a low MTM termination charge.

In our model, the efficient MTM termination charge is (approximately) equal to cost. The reason is that mobile subscriber numbers are boosted when the *surplus* from MTM calls is large, so that network externalities are achieved when the MTM call charge is equal cost. On the other hand, it is FTM termination *profit* which boosts subscriber numbers—see expression (32) in the appendix. This explains the different regulatory choices for FTM and MTM termination charges. Indeed, this analysis provides our only significant example of differential welfare-maximizing FTM and MTM charges, assuming such differential charges are feasible.

3.4 Substitution between FTM and (on-net) MTM calls

As discussed in section 1, our assumption that MTM and FTM calls are separate markets is questionable.[55] Given the average call charges in Table 2, we expect that mobile subscribers will typically call others on the same mobile network by using their mobile phone rather than a fixed phone (even when they are at home or the office, and a fixed phone is available). On other hand, it is less plausible at these prevailing prices that off-net mobile calls will be

[55] See Hausman and Wright (2006) for an alternative analysis of this issue. They assume competition between identical mobile operators and focus on an equilibrium in which there is no off-net and on-net call charge differentials. These differentials play a fundamental role in the analysis presented here. On the other hand, they allow for FTM and MTM calls to be imperfect substitutes, which leads to different implications as noted below.

preferred to FTM calls when callers have access to a fixed phone.[56]

However, the fact that people will substitute on-net MTM calls for FTM calls whenever the former's price is lower does not necessary mitigate a mobile network's incentive to set high FTM termination charges. For instance, suppose a mobile network does set a high FTM termination charge. Then, by means of an appropriate choice of on-net call charge, this network can determine whether its subscribers use FTM calls or on-net MTM calls to call others on the same network. If the network sets a higher on-net call charge than the FTM call charge, its subscribers will prefer to use their fixed phones where feasible, and the network will enjoy the high profit from FTM call termination. The joint surplus available to the mobile network and its subscriber is then

$$v(P(A)) + F(A)$$

with this arrangement.[57] If assumption (1) holds, this surplus is

$$v(P) + q(P)(P - C - c_T) \ .$$

If the network sets a lower on-net call charge, its subscribers will prefer to use their mobile phones, in which case its subscribers benefit from the low call charge. The joint surplus is then

$$v(p) + q(p)(p - c_0 + c_T) \ .$$

If the costs of the two call types are not too dissimilar ($C \approx c_0$), then when A is set above cost c_T, more joint surplus is generated by the second method. Indeed, in that case, the maximum surplus is achieved by setting the on-net call charge equal to cost: $p = c_0 + c_T$. In sum, whenever the origination costs are not too dissimilar for the two types of call, we expect that a mobile network will be able to extract more profit from its subscribers (for a given level of subscriber surplus) by undercutting the FTM call charge with its on-net MTM call charge, even though the network makes less profit from FTM call termination. This is due to the usual dead-weight losses associated with above-cost pricing of FTM calls. Notice that this argument is valid even if it is somewhat less efficient to make on-net MTM calls than FTM calls (so $C < c_0$). In this case, when the FTM termination charge is well above cost, a mobile firm will choose to encourage its subscribers to use the less efficient MTM mode of communication. For instance, with the call charges in Table 2, it is plausible that mobile subscribers make too many on-net calls relative to FTM calls.[58]

[56] This discussion is overly simplified, since the average call charges in Table 2 mask the fact that many individual off-net calls are priced below FTM calls, and many on-net calls are more expensive. For instance, off-net calls are often included in a network's package of free minutes in subscription plans.

[57] Here, we use the same notation as in section 2. However, since here the two kinds of call types are assumed to be (perfect) substitutes when both types are feasible (i.e., when the mobile subscriber has access to a fixed phone), we use the same demand function $q(\cdot)$ for both types of call. (In section 2 we used $Q(\cdot)$ to denote the demand for FTM calls.)

[58] In the enquiry (Competition Commission, 2003, paras. 2.408 and 2.424), BT raised this issue that high FTM termination charges acted to distort competition between FTM and on-net MTM calls.

However, the mobile network (network i, say) has a strict incentive to set its FTM termination charge above cost for the standard competitive bottleneck reason, so as to provide its customers with the most attractive packages (funded by the tax on fixed-line callers who are not subscribers to i's network). Similarly, when they have access to a fixed phone, subscribers to rival network j will prefer to make calls to i's subscribers via a fixed phone, if the off-net MTM call charge is above the FTM call charge. Whether the off-net MTM call charge is above the FTM call charge (as in Table 2) will depend on the prevailing MTM termination charge. In the case where networks set the same FTM and MTM termination charge (as in the UK at the time of the enquiry), and when MTM and FTM calls are viewed as perfect substitutes, the result will be that off-net MTM calls do not significantly substitute for FTM calls. Given the prices prevailing in the UK at the time of the enquiry, we expect off-net MTM calls were not significantly substituting for FTM calls.[59]

In sum, when the call types are perfectly substitutable a network can insulate its own subscribers from the effects of its choice of a high FTM termination charge. It can do this by setting a low on-net MTM call charge (equal to cost, say). The only way one of its subscribers could be affected by its high FTM termination charge is if he wished to call someone else on the same mobile network, and this effect is eliminated when the on-net call charge is low. In such an environment, it seems plausible that a mobile network has a similar incentive to set a high FTM termination as was the case when the two types of calls were not substitutable, and the "competitive bottleneck" problem we emphasized before is not overturned by this substitutability.

Nevertheless, the welfare costs of setting FTM termination charges too high will be somewhat reduced by substitutability. For instance, consider a situation with four equally-sized mobile networks and complete mobile penetration (so all fixed users also have a mobile phone). Then, compared to the situation when there is no such substitution, the volume of FTM calls will be reduced by 25% and so are the corresponding profits from FTM termination. Even though this may have no impact on the equilibrium unregulated choice of FTM termination charge (as argued above), it does reduce the consequent welfare loss by 25%. Similarly, if there are market expansion possibilities (as discussed in section 3.3), then above-cost pricing of FTM call termination may have less impact on market expansion. The profit from FTM call termination is reduced by 25% because of substitution, and so a given FTM termination charge generates a correspondingly reduced subsidy for mobile subscribers.[60]

[59] In other situations, such as the case of Australia studied by Hausman and Wright, even if FTM and MTM termination charges are set at the same level, there may be significant substitution between the two types of calls. This can reflect high FTM prices where FTM calls are not regulated and the offering of some free or discounted off-net minutes by mobile operators in their subscription plans. In addition, it may be that consumers view the two types of calls as imperfect substitutes, so that at equal prices they will make some FTM and some MTM calls. Hausman and Wright adopt this approach, in which case FTM termination charges will be constrained by both on-net and off-net MTM substitution (constrained by on-net substitution since an operator will no longer want to increase FTM termination charges as much given this will force its own subscribers to substitute to MTM calls which the subscriber may not like).

[60] However, when mobile expansion is possible, high FTM call charges will also act to induce people to

3.5 Countervailing power by fixed-line operator

A running theme in this paper concerns the differences in the incentive to choose FTM and MTM termination charges. One difference we have highlighted is that FTM termination charges are often set unilaterally while MTM termination charges might be set by negotiation between mobile networks. In section 3.2 we argued that if MTM termination charges are also set unilaterally, the way they will be set is then not so different from FTM termination charges. Likewise, in this section we show that if FTM termination charges are set by negotiation with the fixed-line network (which can also refuse to deliver calls to a mobile network), then the equilibrium FTM termination charge depends on relative bargaining power. In particular, when a powerful fixed network can refuse to interconnect with a mobile network, it is no longer the case that FTM termination charges are necessarily set above the efficient level (just as with MTM termination charges).

In the analysis so far, the fixed-line operator was assumed to be obliged to interconnect with any mobile operator at the latter's chosen FTM termination charge. Consider now the opposite case in which there is no such obligation, so that the fixed-line operator can refuse to send calls originating on its network to a mobile network if it is not in its interest to do so.[61] For simplicity, suppose there is a single fixed network. In stage one, two mobile networks simultaneously choose their FTM termination charge. Given these choices, in stage two the fixed network chooses with which mobile operator(s) to interconnect, and in stage three the mobile networks set their retail tariffs and compete for subscribers. Crucially, we assume there are call externalities on the part of mobile subscribers, in that, all else equal, they prefer joining a mobile network on which they can receive calls from the fixed network.[62]

Suppose that if the fixed network pays the FTM termination charge A_i to network $i = 1, 2$, it sets the FTM call charge to that network equal to $P(A_i)$. Suppose also that the fixed network enjoys a positive margin on FTM calls, so that $P(A_i) > A_i + C$. (For instance, regulation of the fixed network may be imperfect, or the fixed network is permitted to include a contribution to its fixed costs in its FTM call charge.) Moreover, suppose that the fixed network strictly prefers to pay a smaller termination charge.[63]

Compared to the previous model we make two simplifying assumptions. First, we assume mobile operators are homogenous price competitors. Second, we abstract from MTM calls, focusing instead on FTM calls. Using the notation of section 3.1, the utility of a subscriber

subscribe to a mobile network (to avoid high FTM prices). This introduces new welfare effects, which Hausman and Wright (2006) explore.

[61] See Binmore and Harbord (2005) for a more detailed treatment of this issue which focusses on the bargaining process between a large fixed network and a small mobile entrant.

[62] We assume that the fixed network's own termination charge continues to be regulated, and any mobile network can send calls from its subscribers to the fixed network at this charge. If, in addition to being able to refuse to send its calls to a mobile network, the fixed network could also refuse to deliver calls from the mobile network to its own subscribers, then the fixed network's bargaining power is even greater.

[63] That is to say, its FTM profit $Q(P(A))(P(A) - A - C)$ is decreasing in A.

on network i is

$$u_i = U + BQ(P(A_i)) - r_i$$

if that network is interconnected with the fixed network, while $u_i = U - r_i$ if the network cannot receive calls from the fixed network. (Here, U is the utility subscribers enjoy from making and receiving MTM calls and making MTF calls, which is assumed constant in the following discussion.) Subscribers will all join the mobile network which offers the higher value of u_i.

Consider the outcome if only one mobile firm, say firm i, has secured interconnection with the fixed operator. The best that firm j can do is to offer its restricted service at cost, so $r_j = f$ and $u_j = U - f$. Provided this covers its costs, firm i will set the rental fee which just attracts subscribers, so that $r_i = f + BQ(P(A_i))$. This rental fee covers i's costs provided $BQ(P(A_i)) + \pi_T(A_i) \geq 0$, i.e., provided

$$A_i \geq c_T - B \ . \tag{16}$$

In this case, the mobile firm which has exclusively secured interconnection with the fixed network will serve the entire mobile market.

This insight tells us that in stage 2, if the two mobile networks offer different termination charges (which satisfy (16)), the fixed operator will choose to interconnect only with the network which offers the lower charge. If it does so, the mobile firm with the lower termination charge will serve the entire mobile market, which means that the fixed network will pay the lower charge on all its FTM calls. Therefore, in stage one, each mobile network will choose A_i as low as possible subject to covering its costs.

In essence, the mobile networks compete in a winner-take-all fashion for the right to deliver the fixed operator's FTM calls. The outcome of this contest is that both mobile firms offer to terminate FTM calls at the lowest level which satisfies (16), which is the termination charge $\hat{A}_W$ in expression (13). The fixed network is happy to interconnect with both operators when they offer the same termination charge. Equilibrium rentals are then $r_1 = r_2 = f + BQ(P(\hat{A}_W))$. Therefore, in contrast to the case where the fixed operator was obliged to interconnect with mobile operators, now it is the fixed-line callers who are subsidized by mobile subscribers. The equilibrium FTM termination charge is set below cost. The resulting equilibrium FTM price is some markup over $C + c_T - B$, compared to the socially efficient FTM price (12). Thus, the equilibrium unregulated outcome in this case still results in a socially excessive FTM call charge.

However, if the fixed network's markup is not substantial, so that $P(A) \approx A+C$, then the unregulated outcome corresponds closely to the ideal regulated outcome. This suggests that in some circumstances, an alternative remedy to direct regulation of FTM call termination would be to permit a dominant fixed network to refuse to interconnect with a mobile network.

This discussion assumes there is no regulatory intervention in the setting of FTM termination charges. Even if the parties cannot agree, the authorities will not intervene. With a single fixed network and multiple mobile operators, the natural outcome is that the fixed network holds all the bargaining power. In our previous analysis of the basic model, the authority was assumed only to intervene on behalf of the mobile operator, forcing the fixed network to interconnect and pay the mobile operators' termination charge, thereby giving mobile operators all the bargaining power. A more realistic case lies between these two extremes, in which the authority intervenes when the parties cannot reach an agreement and imposes its own preferred termination charge in such cases, which would then lead to a different outcome (presumably close to the welfare-maximizing choice).

The Commission did not appear to consider that countervailing buyer power might act as a constraint on the mobile networks' ability to set high FTM termination charges. However, this issue has recently emerged, with disputes about whether the new mobile network, *H3G*, should be exempt from regulatory control of its termination charges.[64]

4 Conclusions

The Commission concluded that the four mobile networks should each be subject to two termination charge caps, set at the same level, one to control FTM termination charges and one to control MTM termination charges.[65] The cap on average termination charges (separate for the two kinds of termination) for the combined 900/1800 MHz spectrum networks (i.e., *Vodafone* and O_2) was set at 9 pence per minute for 2002/03 and falling (sharply) to 4.7 pence per minute in 2005/6, all measured in 2002/3 prices. In particular, networks were to be permitted to set different termination charges for FTM and MTM traffic, but these would each be controlled.

The Commission's decision to regulate both termination charges seems consistent with the economic models we have presented. While there is a coherent economic rationale for setting the FTM termination charge above cost in order to boost mobile penetration rates, it is probably also true that the benefits of subsidizing mobile subscribers with inflated FTM call charges decline as mobile penetration levels off. The downward path of termination charges prescribed by the Commission is consistent with this view.

While the theoretical models we presented indicate a potential danger of mobile networks agreeing to set MTM termination charges which are inefficiently low, such models were rather intricate and assumed that mobile firms negotiate agreements with each other to set these low termination charges. Indeed, other situations (such as our model of unilaterally chosen charges in section 3.2, or issues concerning the impact on potential future entry) point to a danger of inefficiently high MTM termination charges. It appears that the Commission put more weight on the latter consideration than the former, and placed a ceiling, but not a floor, on MTM charges. On the other hand, their decision to make the two kinds of termination operate under separate caps indicates that the Commission did not want to see potentially

[64] See Ofcom (2006, paras. 4.23-4.87) for further discussion.

[65] See Competition Commission (2003, para. 2.578 and Table 2.12). The four networks were treated roughly symmetrically, with some small differences to reflect differences in spectrum allocation.

low MTM termination charges lead to higher FTM termination charges. After the event, it appears their judgement was correct: since the report the mobile networks have set a uniform termination charge for both FTM and MTM traffic, and there has been no evidence that they have attempted to negotiate low MTM charges.

After the Commission's judgement (in January 2003), several of the mobile networks appealed against the ruling. However, the High Court rejected the appeals in June 2003 and the Commission's judgement was implemented. Soon afterwards, European policy required Ofcom (the new UK combined telecommunications and broadcasting regulation set up in 2003) to revisit policy towards mobile call termination. Ofcom conducted a review of the voice call termination market in 2003/4 and concluded in June 2004 that regulation of the form suggested by the Commission was still needed.

The sector continues to be highly controversial. The fifth mobile, *H3G*, had entered the market before the June 2004 statement by Ofcom, but Ofcom did not choose to regulate its termination charges (although Ofcom did determine that *H3G* did have significant market power in call termination). *H3G* has since claimed in appeal that it does not have significant market power, arguing that BT has "countervailing buyer power" (as in our stylized model in section 3.5). (However, Table 5 above suggests that *H3G* is nevertheless able to set very high termination charges.) This issue is legally complex and still unresolved.[66] In addition, regulation currently only applies to voice services, and termination charges for text (SMS) services are not controlled. Most of the issues discussed in this paper also apply to text messages, which are an increasingly important and profitable part of the mobile market, and it may be necessary in the future to examine the efficiency of termination charges for these newer services.[67]

Technical Appendix

A Call Externalities and Price Discrimination

Suppose the market share of network i is

$$s_i = \frac{1}{K} + \frac{u_i - \bar{u}_i}{2t} \tag{17}$$

where

$$\bar{u}_i = \frac{1}{K-1} \sum_{j \neq i} u_j$$

[66] See Ofcom (2006b) for more detail.

[67] SMS termination charges are already controlled in France and Israel, for instance. Ofcom has announced that it plans to review the market for SMS termination in 2007/8—see Ofcom, *Wholsale SMS Termination Market Review*, 13 September 2006.

is the average of the other networks' offered utilities. (This demand system reduces to the standard Hotelling specification (3) when $K = 2$.) Again, the total market size is exogenously fixed (and normalised to 1), and firms share the market equally if they offer the same utilities.

To calculate the equilibrium charges given an initial choice for the MTM termination charge a, suppose that each of $K-1$ networks choose the same charges $(\hat{p}, p, r, A)$, and the remaining network i is considering its own charges $(\hat{p}_i, p_i, r_i, A_i)$. A consumer's utility from subscribing to network i changes from (2) to become

$$u_i = s_i[v\left(\hat{p}_i\right) + bq(\hat{p}_i)] + (1-s_i)[v\left(p_i\right) + bq(p)] + BQ(P(A_i)) - r_i \ ,$$

where p is the off-net price set by each of the other networks. This expression includes several components. The utility from outbound calls is $s_i v\left(\hat{p}_i\right) + (1-s_i)\, v\left(p_i\right)$, where this is made up of on-net and off-net calls. Utility is also obtained from receiving calls from on-net MTM callers $bs_i q(p_i)$, from off-net MTM callers $b\left(1-s_i\right) q(p)$, and from fixed line callers $BQ(P(A_i))$.

Similarly, a consumer's utility at each of the other networks is

$$\bar{u}_i = \frac{1-s_i}{K-1}[v\left(\hat{p}\right)+bq(\hat{p})]+s_i[v\left(p\right)+bq(p_i)]+\left[1-s_i-\frac{1-s_i}{K-1}\right][v\left(p\right)+bq(p)]+BQ(P(A))-r \ .$$

Firm i's profit is

$$\begin{aligned}\pi_i = s_i \times [&s_i\left(\hat{p}_i - c_O - c_T\right) q(\hat{p}_i) + (1-s_i)(p_i - c_O - a)q(p_i) \\ &+(1-s_i)\left(a-c_T\right) q(p) + r_i - f + F\left(A_i\right)] \ .\end{aligned} \tag{18}$$

The trick to solving this complicated interaction is to notice that network i's market share is unaffected if it modifies its charges $(\hat{p}_i, p_i, r_i, A_i)$ in such a way that the utility difference $u_i - \bar{u}_i$ is unchanged. In a symmetric equilibrium we have $s_i = \frac{1}{K}$. Therefore, in a symmetric equilibrium, firm i can vary $\hat{p}_i, p_i$ and A_i without affecting its market share $s_i = \frac{1}{K}$, provided that its fixed rental charge satisfies

$$r_i = \frac{1}{K}v(\hat{p}_i) + \frac{K-1}{K}v(p_i) + b(\frac{1}{K}q(\hat{p}_i) - \frac{1}{K}q(p_i)) + BQ(P(A_i)) + \text{ constant} \ . \tag{19}$$

First, consider network i's incentive to choose its FTM termination charge, A_i.[68] Substituting (19) into network i's profit in (18) shows that the unregulated equilibrium FTM termination charge, denoted $\hat{A}_M$, will be chosen to maximise (11) in the text.

Consider next the choice of MTM termination charge. Suppose that mobile networks each set the FTM termination charge A. (As discussed, without regulation we expect firms will choose $A = \hat{A}_M$ but with regulation we may have the efficient charge $A = \hat{A}_W$ or some

[68] The results obtained here are similar to those in Armstrong (2002) and Wright (2002), who considered the impact of call externalities on FTM termination charges.

other charge.) Substituting (19) into (18) shows that each network will set the on-net call charge $\hat{p}$ and off-net call charge p as given by (14) in the text.

Having determined the equilibrium call charges, we complete the analysis of second stage decisions by considering the choice of the rental charge. Then (17) implies that firm i's market share s_i satisfies

$$s_i = \frac{1}{K} + \frac{1}{2t}\left[r - r_i + \frac{Ks_i - 1}{K-1}(v(\hat{p}) + bq(\hat{p}) - v(p) - bq(p))\right] .$$

Solving this explicitly implies that

$$s_i = \frac{1}{K} - \frac{r_i - r}{2t - \frac{K}{K-1}(v(\hat{p}) + bq(\hat{p}) - v(p) - bq(p))} . \tag{20}$$

From (18), with the call charges in (14) network i's profit is

$$\pi_i = s_i \times \left[-s_i bq(\hat{p}) + (1 - s_i)(a + \frac{b}{K-1} - c_T)q(p) + r_i - f + F(A)\right] .$$

From (20), in symmetric equilibrium this implies that each network will choose the rental charge given a so that total industry profit in the mobile sector, denoted Π, is given by

$$\Pi = \frac{1}{K}\left(bq(\hat{p}) + (a + \frac{b}{K-1} - c_T)q(p(a)) + 2t - \frac{K}{K-1}(v(\hat{p}) + bq(\hat{p}) - v(p(a)) - bq(p(a)))\right) . \tag{21}$$

(Recall that the off-net price $p(a)$ is given in (14).) As in section 2, the waterbed effect implies that profit from FTM termination, $F(A)$, has no impact on the equilibrium profit (21). However, unlike the previous analysis, mobile networks are no longer indifferent to the choice of MTM termination charge. Without regulation, the industry will wish to choose a to maximize (21), so that

$$a = c_T - \frac{K+1}{K-1}b + \frac{1}{K-1}\frac{q(p)}{q'(p)} .$$

The resulting off-net call charge is given in expression (15) in the text.

B Unilateral Choice of MTM Termination Charges

Return to the situation with two symmetrically placed mobile networks, where the market share of network i is given by expression (3) above. For simplicity, here we ignore the fixed telecommunications sector in this analysis, since it plays no significant role. Similarly, we ignore call externalities.

Suppose network i sets the on-net call charge $\hat{p}_i$, the off-net call charge to its rival p_i, the fixed rental charge r_i and the MTM termination charge a_i. A subscriber's utility at firm i is

$$u_i = s_i v\left(\hat{p}_i\right) + (1-s_i) v\left(p_i\right) - r_i \ .$$

Similarly to expression (18), firm i's profit is

$$s_i \times \left[s_i\left(\hat{p}_i - c_O - c_T\right) q(\hat{p}_i) + (1-s_i)(p_i - c_O - a_j) q(p_i) + (1-s_i)(a_i - c_T) q(p_j) + r_i - f\right] .$$

Similarly to expression (14) above, one can show that call charges are set equal to the firm's marginal costs:

$$\hat{p}_i = c_O + c_T \ ; \ p_i = c_O + a_j \ .$$

In this case, the firm's profit is

$$\pi_i = s_i \times \left[r_i - f + (1-s_i) M(a_i)\right] \ , \tag{22}$$

where

$$M(a) \equiv (a - c_T) q(c_O + a)$$

measures per-subscriber profit from MTM termination.

From (3) firm i's market share satisfies

$$s_i = \frac{1}{2} + \frac{s_i \hat{v} + (1-s_i) v_j - r_i - \left[(1-s_i)\hat{v} + s_i v_i - r_j\right]}{2t}$$

and so

$$s_i = \frac{1}{2} + \frac{r_j - r_i + \frac{1}{2}(v_j - v_i)}{2t + v_1 + v_2 - 2\hat{v}} \ . \tag{23}$$

(Here, $\hat{v} = v(c_O + c_T)$, $v_j \equiv v(c_O + a_j)$ and $v_i \equiv v(c_0 + a_i)$.) Differentiating (22) with respect to r_i yields

$$r_i = f - \frac{s_i}{\frac{\partial s_i}{\partial r_i}} - (1-2s_i) M_i = f + s_i(2t + v_1 + v_2 - 2\hat{v}) - (1-2s_i) M_i \ , \tag{24}$$

where we have written $M_i = M(a_i)$ and the second equality follows from (23). Substituting this value for r_i into (22) shows that firm i's profit is

$$\pi_i = s_i^2 (2t + v_1 + v_2 - 2\hat{v} + M_i) \ . \tag{25}$$

Moreover, substituting r_i as given in (24) and the corresponding expression for r_j into (23) shows that

$$s_i = \frac{1}{2} + \frac{1}{2}\left(\frac{v_j - v_i}{6t + 3(v_1 + v_2) - 6\hat{v} + 2(M_1 + M_2)}\right) \ , \tag{26}$$

which gives the equilibrium market share purely in terms of the two termination charges, a_1 and a_2. Substituting (26) into (25), differentiating with respect to a_i and imposing symmetry ($a_1 = a_2 = a$) implies the following first-order condition for the unilaterally chosen termination charge a is

$$\left.\frac{\partial \pi_i}{\partial a_i}\right|_{a_i=a_j=a} = \frac{1}{4}\left(M'(a) - q(c_O + a)\right) + \frac{q(c_O + a)}{2}\left(\frac{2t + 2v(c_O + a) - 2\hat{v} + M(a)}{6t + 6v(c_O + a) - 6\hat{v} + 4M(a)}\right) = 0 \ .$$

It follows that the equilibrium termination charge satisfies

$$M'(a) = q(c_O + a)\frac{t + v(c_O + a) - \hat{v} + M(a)}{3t + 3v(c_O + a) - 3\hat{v} + 2M(a)} \ . \tag{27}$$

This termination charge is below the monopoly level where $M'(a) = 0$ (in contrast to the FTM charge), but above the efficient level (where $a = c_T$). When t is very large, (27) shows that $M'(a) \approx \frac{1}{3}q(c_O + a)$, which is well above the efficient level, but still lower than the monopoly level. On the other hand, if one considers the extension of this model to more than two mobile networks (using the utility specification (17) above), it appears that the unilateral choice of MTM termination charges converges to the monopoly level ($M'(a) = 0$) as the number of firms becomes large.

C Network Externalities and Market Expansion

Here, suppose there are two networks, no call externalities and no price discrimination (i.e., each firm's off-net and on-net call charges are equal). Suppose also that the MTM termination charge is not chosen unilaterally, but jointly by the two mobile networks. If networks 1 and 2 offer subscribers the respective utilities u_1 and u_2, suppose that network i attracts $n_i = \phi(u_i, u_j)$ subscribers. Unlike the specification in (3), it is no longer the case that the total number of mobile subscribers, denoted $N \equiv n_1 + n_2$, is constant. Suppose network i offers the two-part tariffs with MTM call charge p_i and fixed rental charge r_i. Since mobile subscribers have more people they can call when the mobile market expands, subscriber utility at firm i is modified from (2) to be

$$u_i = Nv(p_i) - r_i \ . \tag{28}$$

Firm i's profit is modified from (5) to be

$$\pi_i = n_i \times [r_i - f + (p_i - c_O - s_i c_T - s_j a)Nq(p_i) + n_j(a - c_T)q(p_j) + F(A_i)] \ ,$$

where $s_i = n_i/N$ is the firm's market share. As usual, it is immediate that each firm will set its FTM termination charge A_i to maximise its profits from termination, $F(\cdot)$. The potential for market expansion has no impact on a mobile network's incentive to set a high FTM termination charge, since even with a fixed market size the termination charge was set at its profit-maximizing level.

When the MTM termination charge is a, the call charge reflects a network's average cost of making a call, so each firm's MTM call charge p is still given by (7). When firms set the same FTM termination charge, equal to A say, and the same MTM call charge $p(a)$, firm i's profit is

$$\pi_i = n_i \times [r_i - f + NS(a) + F(A)] \ ,$$

where $S(a)$ is given in (8). Therefore, in symmetric equilibrium firm i will choose its rental charge r_i to satisfy

$$\frac{\partial n_i}{\partial r_i}[r - f + NS + F] + \frac{1}{2}N\left[1 + \frac{\partial N}{\partial r_i}S\right] = 0 \ . \tag{29}$$

In order to make further progress, it seems necessary to specify subscriber demand ϕ explicitly. To that end, suppose the representation of subscriber demand in (3) is modified in the following way: if the two mobile networks offer utilities u_1 and u_2 then firm i attracts

$$n_i = \frac{1}{2} + \frac{u_i - u_j}{2t} + \lambda u_i \tag{30}$$

subscribers. (This model of consumer demand is sometimes known as the "Hotelling model with hinterlands".) Here, $\lambda \geq 0$ represents the magnitude of the market expansion possibilities. Combining (28) with (30) shows that

$$N = \frac{1 - \lambda(r_1 + r_2)}{1 - 2\lambda v(p)} \ ; \quad n_i = \frac{N}{2} + \left(\frac{1}{2t} + \frac{1}{2}\lambda\right)(r_j - r_i) \ .$$

In what follows we will assume λ is small, in the sense that terms in λ^2 and higher are ignored. (The more general analysis seems to be rather intractable and not illuminating.) Thus, we focus on the "almost saturated market" case. When λ is small, expression (29) shows that the symmetric equilibrium rental charge r is given by the tedious approximation

$$r \approx (f + t - S - F) + \lambda(2tv - 2ft + 2fS + 3tS + 2tF - 2vS - 2SF - 4t^2 - 2S^2) \ . \tag{31}$$

More transparently, the equilibrium market size is approximately

$$N \approx 1 + 2\lambda(w + F - t - f) \ , \tag{32}$$

where $w(a) \equiv v(p(a)) + S(a)$ is total surplus per subscriber from MTM calls. Thus, high profits from terminating FTM calls (F large) induce a market expansion. On the other hand, within the MTM segment it is higher *surplus* (w large, not S large) which drives up subscriber numbers.

Finally, total mobile industry profit is approximately

$$\Pi \approx t + \lambda t(4v + 3S + 4F - 6t - 4f) \ . \tag{33}$$

Note that the waterbed effect discussed in section 2.3 is no longer 100%. Profits from FTM termination, F, are retained to some extent by mobile firms, and are not fully passed onto subscribers. That is to say, the mobile industry as a whole now has an incentive to lobby against proposed regulation to bring FTM termination charges down to cost. (The basic model in section 2 did not shed any light on why it was that mobile networks were opposed to industry-wide regulatory intervention to bring down these termination charges.)

Similarly, mobile firms are not indifferent to the MTM termination charge. From (33), industry profit is maximised if a is chosen to maximize $4v(p(a)) + 3S(a)$, i.e., firms will wish to choose the MTM termination charge which results in a MTM call charge which is below cost:

$$p(a) = c_0 + c_T - \frac{1}{3}\frac{q}{-q'} \ .$$

Clearly, then, firms prefer a MTM termination charge which is below cost.

What are the socially efficient termination charges in this framework? When the networks' offered utility is $u_1 = u_2 = u$, aggregate consumer surplus of mobile subscribers is $\lambda u^2 + u$. Since $u = Nv - r$, when λ is small (31) and (32) imply that total mobile consumer surplus is approximately

$$CS_M = \lambda u^2 + u$$

$$\approx w + F - f - t - \lambda(6tv + 5tS + 4fw - 4Fw - 3w^2 - F^2 + 2fF + 4tF - f^2 - 4ft - 5t^2) \ .$$

Finally the consumer surplus of fixed line subscribers from calls to mobile subscribers is

$$CS_F = NV \approx (1 + 2\lambda(w + F - t - f))V \ ,$$

where $V = V(P(A))$ is the consumer surplus on the fixed network for each mobile subscriber. Assume that FTM calls are charged at cost, so that expression (1) holds.[69] Then total welfare is obtained by summing Π, CS_M and CS_F. Ignoring constant terms, this total welfare is approximately

$$w + W + \lambda(2Ft + 2Fw - 4fw - 2tw + 2FW - 2fW - 2tW + 2wW - F^2 + 3w^2) \ , \quad (34)$$

where $W = V + F$ is surplus per mobile subscriber generated by FTM calls. Clearly, when there are no market expansion possibilities ($\lambda = 0$), it is optimal to set both the FTM and MTM termination charges equal to cost c_T as in our benchmark model. More generally, notice that the MTM call charge p only enters expression (34) through total surplus w and that the expression is increasing in w whenever λ is reasonably small. We deduce it is still optimal for the MTM termination charge to maximize w, i.e., to set $a = c_T$. On the other hand, when $\lambda > 0$ one can show from (34) that the FTM termination charge is optimally

[69] Assume also that the mobile-to-fixed termination charge is regulated to be equal to the fixed network's cost. Otherwise, the extra profits on the fixed network caused by terminating more traffic when the mobile market expands would need to be considered too.

set above cost, but below the unregulated level. (The precise formula is not illuminating.) In the sum, optimal policy in this context entails:

$$a \approx c_T \ ; \ c_T < A < A_M \ .$$

Potential market expansion appears to provide little reason for a regulator to distort the MTM call charge away from cost. Expression (32) indicates that market size is boosted when surplus w is maximized, i.e., when the MTM termination charge is equal to cost.[70]

References

Armstrong, M. (1998): "Network Interconnection in Telecommunications," *Economic Journal*, 108(448), 545–564.

——— (2002): "The Theory of Access Pricing and Interconnection," in *Handbook of Telecommunications Economics: Volume I*, ed. by M. Cave, S. Majumdar, and I. Vogelsang. North-Holland, Amsterdam.

——— (2006): "Competition in Two-Sided Markets," *Rand Journal of Economics*, forthcoming.

Armstrong, M., and J. Vickers (2001): "Competitive Price Discrimination," *Rand Journal of Economics*, 32(4), 579–605.

Behringer, S. (2006): "Asymmetric Equilibria and Non-Cooperative Access Pricing in Telecommunications," mimeo, Frankfurt.

Berger, U. (2005): "Bill-and-Keep vs. Cost-Based Access Pricing Revisited," *Economics Letters*, 86(1), 107–112.

Binmore, K., and D. Harbord (2005): "Bargaining Over Fixed-To-Mobile Termination Rates: Countervailing Buyer Power as a Constraint on Monopoly Power," *Journal of Competition Law and Economics*, 1(3), 449–472.

Carter, M., and J. Wright (1999): "Interconnection in Network Industries," *Review of Industrial Organization*, 14(1), 1–25.

Cherdron, M. (2002): "Interconnection, Termination-Based Price Discrimination, and Network Competition in a Mature Telecommunications Market," mimeo, Mannheim.

[70] This result may not hold if λ is not "small". Indeed, Proposition 4 in Dessein (2003) suggests that the welfare-maximizing MTM termination charge will usually be above cost. Such a policy makes the mobile networks compete harder, and the resulting higher utility offered to subscribers brings forth further subscribers. However, in the "almost saturated" market we emphasize here, the benefits of setting the MTM termination charge above cost will be limited.

Competition-Commission (2003): *Vodafone, O2, Orange and T-Mobile: Reports on References Under Section 13 of Telecommunications Act 1984 on Charges Made by Vodafone, Orange, O2 and T-Mobile for Terminating Calls Made by Fixed and Mobile Networks.* HMSO, London.

Dessein, W. (2003): "Network Competition in Nonlinear Pricing," *Rand Journal of Economics*, 34(4), 593–611.

Dixit, A. (1986): "Comparative Statics for Oligopoly," *International Economic Review*, 27(1), 107–122.

Gans, J., and S. King (2000): "Mobile Network Competition, Customer Ignorance and Fixed-to-Mobile Call Prices," *Information Economics and Policy*, 12(4), 301–328.

——— (2001): "Using 'Bill and Keep' Interconnect Agreements to Soften Network Competition," *Economics Letters*, 71(3), 413–420.

Gans, J., S. King, and J. Wright (2005): "Wireless Communications," in *Handbook of Telecommunications: Volume II*, ed. by S. Majumdar, I. Vogelsang, and M. Cave. North-Holland, Amsterdam.

Genakos, C., and T. Valletti (2007): "Testing the 'Waterbed' effect in Mobile Telephony," mimeo.

Hahn, J.-H. (2004): "Network Competition and Interconnection with Heterogeneous Subscribers," *International Journal of Industrial Organization*, 22(1), 611–631.

Hausman, J., and J. Wright (2006): "Two-Sided Markets with Substitution: Mobile Termination Revisited," mimeo.

Höffler, F. (2006): "Mobile Termination and Collusion, Revisited," mimeo, Bonn.

Hoernig, S. (2007): "On-net and Off-net Pricing on Asymmetric Telecommunications Networks," *Information Economics and Policy*, forthcoming.

Laffont, J.-J., P. Rey, and J. Tirole (1998a): "Network Competition: I. Overview and Nondiscriminatory Pricing," *Rand Journal of Economics*, 29(1), 1–37.

——— (1998b): "Network Competition: II. Price Discrimination," *Rand Journal of Economics*, 29(1), 38–56.

MMC (1999a): *British Telecommunications plc: A Report on a Reference Under Section 13 of Telecommunications Act 1984 for Calls Made from its Subscribers to Phones Connected to the Networks of Cellnet and Vodafone.* HMSO, London.

——— (1999b): *Cellnet and Vodafone: A Report on a Reference Under Section 13 of Telecommunications Act 1984 on the Charges Made by Cellnet and Vodafone for Terminating Calls from Fixed-Line Networks*. HMSO, London.

Ofcom (2006a): *The Communications Market 2006*. Office of Communications, London.

——— (2006b): *Mobile Call Termination: Market Review*. Office of Communications, London.

——— (2007): *Telecommunications Market Data Tables: Q2 2006*. Office of Communications, London.

Valletti, T., and G. Houpis (2005): "Mobile Termination: What is the 'Right' Charge?," *Journal of Regulatory Economics*, 28(3), 235–258.

Wright, J. (2002): "Access Pricing under Competition: An Application to Cellular Networks," *Journal of Industrial Economics*, 50(3), 289–316.