Bremer Energie Institut

Bremen Energy Working Papers

Paperseries No. 03

Gert Brunekreeft and Nele Friedrichsen

Vertical unbundling, the coordination of investment, and network pricing.

August 2010

In Cooperation with Jacobs University Bremen

JACOBS UNIVERSITY

Herausgeber/ Editor:

Bremer Energie Institut

College Ring 2 / Research V

28759 Bremen

Tel. +49 (0) 421 - 200-48 88

Fax +49 (0) 421 - 200-48 77

info@bremer-energie-institut-de

www.bremer-energie-institut.de

Ansprechpartner:

Katy Jahnke

Tel. +49 (0) 421 - 200-48 91

Fax +49 (0) 421 - 200-48 77

E-mail jahnke@bremer-energie-institut.de

Vorgeschlagene Zitierweise/ Suggested citing:
Autor, (Jahr), "Titel", Bremen Energy Working Papers No. #, Bremer Energie Institut.

Vertical unbundling, the coordination of investment, and network pricing

Gert Brunekreeft, Nele Friedrichsen[*]
Jacobs University Bremen, College Ring 2, 28759 Bremen, Germany
email: g.brunekreeft@jacobs-university.de, n.friedrichsen@jacobs-university.de

July 20, 2010
Preliminary version - please do not quote without permission

Abstract

This paper provides a formal analysis on the investment coordination problem in a vertically separated electricity supply industry, although the analysis may apply also to other network industries. In an electricity system, the investment decisions of network and power plants need to be coordinated. In unbundled markets, firm-internal coordination no longer applies. We develop a formal approach to examine whether simple information exchange ("cheap talk") could restore coordination. We adopt a three-stage profit-optimized investment model, with a (regulated) monopoly network and two asymmetrical Cournot-type generators. To analyse credibility of cheap talk we apply the concept of self-signalling in a game with incomplete information and positive spillovers. We show that cheap talk cannot generally solve the investment coordination problem and as a result separation may actually cause a costly coordination problem. We then examine locational network pricing as a coordination device to internalize the incentive problem.

Keywords: cheap talk, unbundling, game theory, network, investment, coordination

JEL-classification: C72, D23, L22, L51

[*]The authors gratefully acknowledge useful comments by Rolf Künneke and by the participants at the workshop "Vertical Relations in Energy Markets" at the University for Economics in Vienna and the conference Enerday 2010 in Dresden. Financial support from the research council NGInfra within the project UNECOM (www.unecom.de) is gratefully acknowledged.

1 Introduction

Liberalisation, market opening and the introduction of competition in various network industries like electricity, gas, telecommunications and transports, is well established in many countries around the world. To promote and enforce competition, we observe a tendency towards vertical separation of monopolistic parts of these sectors (which are usually the networks) and commercial businesses. There are different variations and names of vertical separation; we follow the recent debate in European energy markets where vertical separation of networks is called "network unbundling" (see European Commission, 2007, p.226). The most extreme form of unbundling is ownership unbundling where vertically integrated companies are forced to divest either the network or the commercial businesses. One of the consequences of the liberalisation is a decentralisation of decisions, including investment decisions. Under ownership unbundling, the decisions are decentralised by law or by definition. In large technical systems (as in network industries) the vertical stages in the production chain are complements and usually the decisions and actions require careful coordination. With vertical separation, firm-internal coordination falls away and should be replaced by external coordination of market forces.

In this paper, we will apply our ideas to the electricity sector, but note that the concepts apply to other network industries as well. In the specific case of an electricity sector, ownership unbundling would mean to separate the network from the power plants.[1] In practice, the problem is getting urgent. In many countries, the sector is in the valley of the investment cycle and faces stubstantial investment needs in both network and generation. Moreover, enforced by climate change policy and scarcity of natural resources, the electricity sector awaits large uncertain changes in technology and fuel mix. In particular, we observe plans for large-scale expansion of offshore wind, solar energy, small-scale decentralised generation, clean-coal and nuclear power. These developments have large impacts on the design and expansion of the high-voltage transmission grids. Exactly this is the root of our problem. The optimal development of the network depends crucially on the location and capacity of the power plants to be connected to the network. However, the network needs to be planned years ahead of the planning of the power plants. Moreover, in the liberalised world an investment plan is commercially strategic information which is held back (cf. Brunekreeft and McDaniel, 2007, p.332/333). Firm internal coordination has fallen away and the question then is, how does the market coordinate the simultaneously optimized investment decisions of the network and the power plants?

If the problem is that the network planner does not know the investment plans of the generators, the obvious answer would be that the network plan-

[1] In practice, one should distinguish the debate on the high-voltage transmission network from the low-voltage distribution network. Although the debates are actually different, we note that the main insights of the paper apply to both types of network unbundling. For convenience, we concentrate on the better known debate on high-voltage transmission networks.

ner can simply ask the generators.[2] The obvious counterargument would be strategic behaviour of generators. In other words generators might have an incentive to lie. The coordination problem may thus be an information problem or an incentive problem or both. In this paper, we examine exactly this problem in a formal model and ask whether simple exchange of information can solve the investment coordination problem. In more formal terms, we apply a sequential-moves, one-way communication approach with incomplete information in a situation with positive spill-overs to examine whether cheap talk, as game theoretical concept for simple, costless communication, is credible and can solve the investment coordination problem. In our model, the pay-offs in the decision-matrix are formally derived from a closed, optimised investment game with two vertically-related production stages (a monopoly network and an asymmetric two-firm Cournot generation duopoly). Our cheap-talk credibility criterion relies on the "self-signalling" concept as defined theoretically by Aumann (1990) and Baliga and Morris (2002).

We find as the good news that the number of cases in which there is a problem is relatively small in the first place; there is not always a coordination problem. Nevertheless, we find as the bad news that for our setting in an unbundled electricity sector (with vertical spill-overs), the criteria of self-signalling are violated and thus cheap talk cannot generally solve the investment coordination problem. Moreover, we show that in these cases the uncoordinated outcome is different and welfare decreasing as compared to the integrated outcome. We then continue to examine whether cost-reflective locational network pricing ("deep charging") can internalize the spill-over and restore credibility of cheap talk. We conclude that cost-reflective deep charging helps, but unfortunately, not perfectly so.

This paper is organized as follows. Section 2 briefly reviews the relevant background literature. Section 3 outlines the model and section 4 brings the results of the model. Finally, section 5 gives a discussion of the results and concluding remarks.

2 Literature

2.1 The investment problem in the liberalised electricity supply industry (ESI)

Electricity market organization changed dramatically over the last decades. The incumbent vertically integrated monopolies have been restructured in many countries to foster competition. Vertical network unbundling of the energy sector was fiercely debated by the European Commission in 2008/9. Unbundling

[2] The European Transmission System Operators (ETSO) propose exchange of detailed information about projected time schedules, exact location or connecting point as well as project plan and electrical configuration between investors and grid operators at an early stage of planning as necessary tool to mitigate difficulties in network planning caused by uncertainties (ETSO, 2003, p.19).

of monopolistic networks from generation and supply activities has been introduced to level the playing field for new competitors and prevent discriminatory behaviour by network companies. An important argument brought forward in the debate in favour of unbundling is "strategic investment withholding" (European Commission, 2007; Balmert and Brunekreeft, 2010). Assume areas A and B. A is a high price area and B a low price area and there is a network constraint from B to A. Assume further a local vertically integrated utility with (local) monopoly power in A. Building a larger line to B increases import capacity into A. This would allow third parties from B to supply demand in A and would thus intensify competition for the local supplier in A. Hence, in order to protect its position in generation, the integrated transmission system operator in A will have insufficient incentives to expand the interconnector capacity. Therefore we call this "strategic investment withholding". In contrast, an unbundled transmission system operator does not have any incentive to protect a local generation monopoly (as it does not gain directly from generation profits) and will therefore have stronger incentives to invest in interconnector capacity.

On the other hand, network unbunding decentralizes the decisions in the vertical production chain. Electricity supply is a complex and highly interrelated system, which needs careful coordination (Joskow and Schmalensee, 1983). Any required network upgrading critically depends on generation expansion, which creates a coordination problem. To illustrate this problem, we follow Baldick and Kahn (1993). Consider figure 1. file

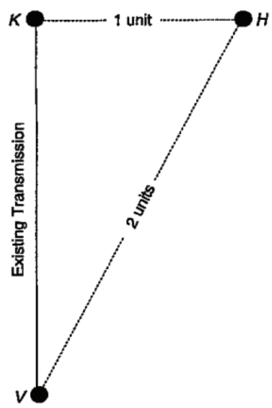

Figure 1: Coordination problem in generation and transmission expansion (Baldick and Kahn 1993, p. 373)

K and H are generation nodes and V is a load node. The line between K and V already exists, but should possibly be upgraded, whereas the other lines still need to be constructed. If the generation capacity of K or H or both is expanded, the network needs to be upgraded. Baldick and Kahn (1993) distinguish two options. First, 'radial connections', which refers to a direct connection between generator and load (K-V and H-V), and second, 'network connection', which implies building the line H-K, such that the connection between H and V is in fact the combination H-K-V. Network expansion takes a longer route (than radial connection H-V) but can be cheaper, for instance, due to economies of scale in the K-V expansion (Baldick and Kahn, 1993, p.374). Baldick and Kahn (1993) illustrate how optimal transmission investment depends on the division of generation capacity expansion among the two generators (at K and H). Radial expansion tends to be optimal if either of the generators takes over most of the additional generator capacity. If, on the other hand, the additional capacity were divided more or less evenly over the two generators, network expansion would be optimal. The coordination problem arises with new investment at H, while it is unknown what will happen at K. The important lesson is that the lack of information about investment plans of third party investors complicates network planning and possibly leads to inefficiencies. We argue that unbundling contributes to the problem because it eliminates firm internal coordination of generation and network. Empirical studies confirm diseconomies of vertical separation in electricity networks (e.g. Nemoto and Goto, 2004; Kwoka, 2002), but the specification of the effects often remains vague. We focus on inefficiencies resulting from a lack of coordination and refer to them as "coordination costs" or "transaction costs".

The "locational" problem illustrated above is currently highly relevant because enormous amounts of new generation plants are planned to be constructed. Consider the example of wind and coal power both locating in near the coast[3] which requries transmission networks to be reinforced to transport the power to the load centers that are typically not located in coastal areas (see e.g. for Germany DENA (2005) and for The Netherlands TenneT (2009). The problem is that investors in generation capacity mostly do not pay for the costs of network reinforcement they cause. Usually, network costs are socialized to the end-users, instead of being cost-reflective and sending locational signals to users. Similar problems are meanwhile arising in distribution networks. In Great Britain a relatively high investment need is foreseen for distribution networks due to tight capacity, growing demand, and the move to a low carbon economy causing increasing amounts of decentralized generation (DG) being connected (cf. OFGEM, 2009). This imposes new challenges on the networks. Efficiency requires some degree of coordination between network and generation investments, because the effects on the network from new DG depend significantly on location apart from technology, network design, and operation (cf. Ackermann, 2004; Prica and Ilic, 2007). Note that effects can be either positive or negative.

2.2 Coordination problems and cheap talk

Cheap talk as game theoretical concept describes communication between players that does not directly influence payoffs. Cheap talk is neither costly nor binding and players may tell the truth or lie, and may or may not believe each

[3]Wind power increasingly moves offshore. Attractivity of the North for coal plants results from cheap transportation cost and better availability of sites.

other (cf. Aumann and Hart, 2003, p.1619). We use cheap talk as a formal model of information exchange as a coordinating device. We are interested in the cases where cheap talk can effectively solve the coordination problem. This requires some information be conveyed about the intended action (Aumann, 1990). As cheap talk is a costless information signal, credibility of the signal is not backed-up by the cost of the signal. Whether or not information can be transmitted and cheap talk is credible depends on the structure of the problem. The credibility of cheap talk has been extensively studied in the literature. We refer the interested reader to Farell and Rabin (1996) for a good overview. The literature distinguishes two credibility criteria. First, *self-committing* (Farell, 1988) and second, *self-signalling* (Aumann, 1990). Farell (1988, p.212) suggested that a message were credible "if the suggested move be rationalizable when others are expected to follow the suggestion". This has been referred to as self-committing cheap talk: the expectation that the cheap talk statement is believed creates an incentive to act according to the signal (cf. Baliga and Morris, 2002). This concept has been challenged as being insufficient for credibility because if a player has strict preferences over the others' actions he wanted him to take a certain action independent of what the own intended action is. Hence, the signal does not convey any information (cf. Aumann, 1990, p.616). Following Aumann (1990) *credible* cheap talk conveys information about the desired behaviour of the opponent *and* about the own intended action. Hence, by signalling to play a certain equilibrium and wanting the message to be believed, the sender is also signalling that he intends to stick to his signal. This condition for credibility of cheap talk is known as *self-signalling*: the sender would want the signal to be believed if and only if it was true (cf. Farell, 1993; Baliga and Morris, 2002).

Coordination via cheap talk has been studied in different setups among others by Crawford and Sobel (1982); Baliga and Morris (2002); Aumann and Hart (2003). An important outcome of prior work has been that positive effects from cheap talk are more likely the closer players' preferences (cf. Crawford and Sobel, 1982). Baliga and Morris (2002) study coordination via cheap talk under existence of spillovers in a two player incomplete information game. A positive spillover means that an investment by player 2 also benefits player 1. Baliga and Morris (2002, p.457) claim that self-signalling is the stronger credibility criterion and argue that "with incomplete information" [...] "the need for self-signalling and the incentive problems created by positive spillovers emerge naturally from the equilibrium analysis." Indeed, the self-committing concept as in Farell (1988) applies to complete information in simulatenous games where signalling would then be the way out of a genuine coordination problem (like the battle of the sexes). The game developed by Baliga and Morris (2002) relies on incomplete information and one-way communication. Whereas player 1 has all the information on the pay-off structure, player 2 does not know the characteristics of player 1. Therefore, player 2 must rely on the signals from player 1. As Baliga and Morris (2002) show, there can be situations where player 1 may lie to trick player 2 into an action it would not choose if player 1 were truthful. This would be a violation of the self-signalling conditions, following definition 3 in Baliga and Morris (2002, p.455). If this is the case, even for only a small subset of all outcomes, cheap talk breaks down generally, because player 2, who does not know the payoff structure, can never know whether player 1 is lying or not.

The situation we study echoes the underlying assumptions in Baliga and Morris (2002), in particular incomplete information, one-sided signalling, and positive spill-overs. We model the situation where the network investor must decide on expanding the network or not, depending on power plant capacities (or locations). The network investor must (irreversibly) invest ahead of the generators and must thus rely on the signals from the generators. The positive spillover is the assumption that a larger network lowers the production costs of the generators (or, to put it more realistically, lowers the probability of not being able to produce due to a congested network).

2.3 Locational network charging

In cases where simple information exchange cannot achieve coordination, network tariffs could signal users their network impact and thus influence investment decisions and operation. Locational components can be realized in the connection charges by making them deep or shallow. "Shallow" charges allocate only the connection cost to the next grid access point to the user. "Deep" charges require the user to also pay for reinforcements that become necessary deeper in network as a consequence of his connection. While "deep" charges signal network impact to network users, which is considered favourable, implementation is not without problems (cf. Brunekreeft et al., 2005; Scheepers et al., 2007). In practice we observe mostly shallow-ish charging and deep charging only in exceptional cases. In view of the upcoming investment needs in distribution networks, the UK energy regulator, Ofgem, investigated more cost reflective charging approaches. These could enhance economic efficiency by directing users away from congested parts and encourage usage where there is surplus capacity. Several studies commissioned by Ofgem investigated pricing to reflect forward looking cost of network development as well as differences in location (cf. Li et al., 2005; Li, 2007; Strbac and Mutale, 2007). In two steps, in 2010 and 2011, Ofgem now introduces a new common charging methodology for distribution networks that aims to signal required locational network reinforcements and thus avoid expansion where possible (cf. OFGEM, 2009).

Locational signals can also be realized in a spot market where efficient electricity prices would reflect the differences over location and time (cf. Schweppe et al., 1988), i.e. nodal spot pricing or locational marginal pricing. It is argued that locational marginal pricing "takes all relevant generation and transmission costs appropriately into account and hence supports optimal investments" (Hogan (2008, p.12) citing IEA (2007, p.116)). However, due to scale economies in network expansion, it is questionable whether nodal prices alone are sufficient to coordinate generation and transmission efficiently (cf. Brunekreeft et al., 2005). They indicate the right direction but probably insufficiently. A combination of locational network tariffs and zonal pricing used complementary to send both short term and long term signals for coordination of generation and networks might be favourable.

3 The model

For our formal approach we use the following notation.

Strategic players are denoted with the following subscripts:

A	- network owner/investor
B_i	- generators $i = 1$ and 2

The parameters and variables are as follows:

P	- price
P_I	- price for the intermediate product I, (i.e. energy, without the network charge)
E	- end users
μ_E	- network charge payable by end-users E
μ_B	- network charge payable by B_i for $i = 1, 2$
μ	- total network charge $\mu_E + \mu_B$
Q	- output, with Q_A, Q_{B1} and Q_{B2} resp.
K_A^S	- capacity of network A
K_{Bi}^S	- capacity of generator B_i
s	- capacity choices low or high, with $s \in \{L, H\}$; s can be different for Q and K: s_Q and s_K
σ^S	- cheap talk signal by B_1 on planned capacity choice
n	- state of demand, decided by nature $n \in \{L, H\}$
α	- probability of the state of high demand $n = H$ and $(1 - \alpha)$ for low demand $n = L$
H, L	- denote "high" and "low" resp.
c_A, c_B	- short-run marginal cost (on output Q)
β_A, β_B	- long-run marginal costs - i.e. capacity expansion costs
γ_1, γ_2	- network scarcity cost increase factor for B_1 and B_2 respectively (where we dropped the B for ease of notation)
z_1, z_2	- "deep" charges payable by generators B_1 and B_2.

Network expansion and the assumption of asymmetry: the role of γ_1.

The "network scarcity cost increase factor" (γ_1, γ_2) is a critical factor which deserves some attention. As this factor applies to players B only, for convenience, we dropped the subscript B in the notation. We assume that $\gamma_1 = 1$ if $K_A = K_A^H$ and $\gamma_1 = \gamma_2 > 1$ if $K_A = K_A^L$. The simple assumption is the main driver of the model and creates a positive spill-over. It says that if the network is small, it gets congested causing higher costs for network users (the generators); or the other way around, expansion of the network lowers production cost for network users. However, importantly and as explained above, we introduced an asymmetry. The network expansion benefits generator B_1, but not B_2. This assumption has two reasons. First, it is far more realistic for our context of *location* of investment and the subsequent need to adjust the network, which is asymmetric almost per definition. Second, we explicitly model asymmetry as the main driver of our results. In this case, if the network is expanded, B_1 benefits and gains a competitive advantage over B_2. It then depends on cost allocation whether this is profitable for B_1 or not. Further below we discuss the comparison with symmetry.

Our approach is a 3-stage sequential-moves game with one-way communication, where lying is explicitly allowed. We assume three players: a network owner/ investor A, and two generators B_1 and B_2. We assume that the generators at B are asymmetrical in γ,[4] but symmetrical in all other parameters ($c_{B1} = c_{B2} = c_B$ and $\beta_{B1} = \beta_{B2} = \beta_B$). Also, we make a simplifying assumption without loss of insight. We assume for simplicity that all changes and capacity choices are for B_1 but not at B_2, (although B_2 does optimize short-run output Q_2). This means that $\gamma_2 = \overline{\gamma}_2 > 1$ and $K_{B2} = \overline{K}_{B2}$ are constants and are not variables within the control of the firm B_2. If B_2 also chooses capacity, the number of cases in our decision tree would increase from $2^2 = 4$ to $2^3 = 8$, without gaining additional insight. It is sufficient for our purposes that B_1 can gain a competitive advantage over B_2 by a reduced γ_1 if the network A is large.

- Stage 1: B_1 signals its investment decision (its capacity choice): $\sigma^L(K_{B1})$ or $\sigma^H(K_{B1})$.

- Stage 2: using the signal from stage 1, the network owner A, decides irreversibly on the capacity of the network: K_A^L or K_A^H.

- Stage 3: after knowing the capacity of the network, B_1 chooses its capacity K_{B1}^L or K_{B1}^H.

A does not know the payoff structure of B_1 and has to rely on the signal given by B_1. In all stages, we assume (and explicitly model) for all cases that short-run output decisions (Q) are optimized (conditional upon capacity) and are not separate stages in the game. Nature decides on low or high demand at the end of the game after all capacity choices have been made. Therefore, all capacity choices are under uncertainty which is the same for all. We explicitly allow the possibility that B_1 lies; i.e. we allow $\sigma(K_{B1}) \neq K_{B1}$. We solve the problem backwards: first, players optimize output conditional upon capacity, and then capacity choices are made using optimal conditional output choices. Note that capacity choices are binary: either high or low capacity.

We assume complementarity of output (normalized to a one-to-one relation): $Q_E^s = Q_A^s = Q_B^s = Q^s$, where $Q_B^s = Q_{B1}^s + Q_{B2}^s$. We define inverse demand of end-users:

$$P^n(Q) = P_I + \mu_E = a^n - bQ \text{ with } n \in \{L, H\} \tag{1}$$

[4]The asymmetry does not refer to the value of γ, but rather to the fact that γ_1 reduces to 1 for high network capacity, while $\gamma_2 > 1$ persists.

Note that in this formulation of demand, it does not matter whether the network charge is paid directly to the network owner or indirectly to the energy supplier, who subsequently passes it on to the network owner. Note also that there is one price only; and that this price relies on used Q only. Therefore, in principle charges in the network and energy prices have the same effect. There is no multi-part tariff. Uncertainty on demand (low or high) results in a parallel shift of linear demand.

The network owner: A

The cost of the network are formulated as follows:

$$C_A(Q^{s_Q}, K_A^{s_K}) = c_A Q^{s_Q} + \beta_A K_A^{s_K}, \qquad \text{for } s_Q, s_K \in \{L, H\} \text{ and } Q^{s_Q} \leq \min\{K_A^{s_K}, K_B^{s_K}\} \tag{2}$$

and network revenues are:

$$R_A(Q^{s_Q}) = (\mu_B + \mu_E)\, Q^{s_Q} \qquad \text{for } s_Q \in \{L, H\} \text{ and } Q^{s_Q} \leq \min\{K_A^{s_K}, K_B^{s_K}\} \tag{3}$$

We assume that both the total level of network charge is regulated, $\mu_B + \mu_E \leq \overline{\mu}$.[5] The "regulation" reflects a revenue cap or price cap, but we allow non-negative profits: $\overline{\mu} \geq c_A + \beta_A$. This is non-critical, and only simplifies because the network charge is not a variable, which subsequently avoids the effects of double marginalisation.

More interesting is the cost of a network expansion. Here we use what is known as the "used-and-useful" criterion. In other words, the revenue driver for the network is output Q, and not capacity K. Therefore, if network expansion is actually used, it is paid for, but if it is not used, it will not be paid for. Hence the network owner bears the risk. Apart from being a realistic assumption, we need this formulation to set incentives for the network investor to make a decision on network expansion at all. Otherwise the expansion cost would be automatically passed through and it would (almost) always be profitable for the network owner to expand the network, which makes the problem trivial.[6]

The network users: generators B_1 and B_2

We now introduce an approach to "deep" versus "shallow" charging. Define an additional network charge, z, to be paid by B, which may reflect the additional costs of network expansion. z is a regulated (arbitrarily high) number. We maintain the idea that the capacity expansion is financed by the normal charge μ. In contrast, we formulate a "deep charge" such that it only serves as

[5] We might add the split μ_B/μ_E for completeness, but note immediately that by cost-incidence it does not have effect and drops out in our formulation.

[6] The reader may realize that this is the key regulatory problem of "efficient investment". The alternative is full cost-pass-through, which makes the problem of overinvestment worse.

a locational signal and does not contribute to financing the expansion. Therefore, we formulate the deep charge such that it is revenue neutral. Formally, the deep charge is then defined as:

$$z_1 = -z_2 \tag{4}$$

where $z_1 \geq 0$, and $z_2 \leq 0$; and $z_1, z_2 = 0$, for $K_A = K_A^L$; and $z_1 > 0$, $z_2 < 0$ for $K_A = K_A^H$. Obviously, by comparative statics, $z_1, z_2 = 0$ in case of shallow charging. Also note that by the assumed revenue neutrality, z_1, z_2 do not show up in A's profit function.

Define the cost function at B_1 as:

$$C_{B1} = (\gamma_1 c_B + \mu) Q_{B1}^{s_Q} + z_1 + \beta_B K_{B1}^{s_K} \qquad \begin{array}{l} \text{for } s_Q, s_K \in \{L, H\} \\ \text{and } Q^{s_Q} \leq \min\{K_A^{s_K}, K_B^{s_K}\} \end{array} \tag{5}$$

the revenues for B_1:

$$R_{B1} = P^n(Q) Q_{B1}^{s_Q} \text{ for } s_Q \in \{L, H\} \text{ and } Q^{s_Q} \leq \min\{K_A^{s_K}, K_B^{s_K}\} \tag{6}$$

and B_1's profit function:

$$\pi_{B1} = R_{B1} - C_{B1} \tag{7}$$

Functions for B_2 are defined accordingly. Note how the end-users' network charge μ_E automatically drops out, as claimed above. In this formulation the split μ_B/μ_E is irrelevant.

Short-run constrained optimization under 2-firm Cournot

We assume two generators, that are assumed to behave as Cournot competitors. The generators at B optimize output under the Cournot assumption for end-user demand conditional upon μ_B, $K_A^{s_K}$, and $K_B^{s_K}$. Moreover, $Q^{S_Q} \leq \min\{K_A^{s_K}, K_B^{s_K}\}$. Using the usual Cournot optimization, then gives B_1's and B_2's reaction functions:

$$Q_{B1}^* = \frac{1}{2b}\left(a^n - (\gamma_1 c_B + \mu) - bQ_{B2}^*\right) \text{ and} \tag{8}$$

$$Q_{B2}^* = \frac{1}{2b}\left(a^n - (\gamma_2 c_B + \mu) - bQ_{B1}^*\right) \text{ for } n \in \{L, H\} \tag{9}$$

Substituting and solving, then gives optimal equilibrium outputs:

$$Q_{B1}^* = \frac{1}{3b}\left(a^n - (2\gamma_1 - \gamma_2) c_B - \mu\right) \text{ and} \tag{10}$$

$$Q_{B2}^* = \frac{1}{3b}\left(a^n - (2\gamma_2 - \gamma_1) c_B - \mu\right) \text{ for } n \in \{L, H\} \tag{11}$$

Define $Q^* = Q_A^* = Q_{B1}^* + Q_{B2}^*$, which then gives total equilibrium values:

$$Q^* = \frac{1}{3b}\left(2a^n - (\gamma_1 + \gamma_2)\, c_B - 2\mu\right) \text{ for } n \in \{L, H\} \tag{12}$$

and

$$P^* = \frac{1}{3}(a^n + 2\mu + (\gamma_1 + \gamma_2)\, c_B) \text{ for } n \in \{L, H\} \tag{13}$$

The size of capacity

We need to make assumptions on the size of the capacities. These might be exogenously chosen arbitrary numbers, but in order to allow larger generality and express capacity choices in parameter values, we use (optimized) *consistent expectations* for initial capacity choices. In other words, we calculate optimized outputs and then take the relevant cases as inital capacity choices.

First, we assume that $K_{B2} = K_{B2}^L$. Recall that we assumed above that capacity is not a variable for B_2. This assumption then says that, as an arbitrary choice, B_2's capacity is always the optimized capacity of the low demand case.

For the low capacity case, we assume that capacity sizes are:

$$K_A^L = Q_A^{L*},\ K_{B1}^L = Q_{B1}^{L*} \text{ and } K_{B2} = Q_{B2}^{L*} \tag{14}$$

where outputs are optimized outputs and $Q_A^{L*} = Q_{B1}^{L*} + Q_{B2}^{L*}$. Note that due to K_A^L, the values of $K_{B1}^L = Q_{B1}^{L*}$ and $K_{B2} = Q_{B2}^{L*}$ rely on $\gamma_1 > 1$.

For the high capacity case, we assume that capacity sizes are:

$$K_A^H = Q_A^{H*},\ K_{B1}^H = Q_{B1}^{H*} \text{ and } K_{B2} = Q_{B2}^{L*} \tag{15}$$

where outputs are optimized outputs and $Q_A^{H*} = Q_{B1}^{H*} + Q_{B2}^{L*}$. Here, because network capacity is high, $\gamma_1 = 1$.[7]

Rationing

In some cases we find by mechanism that optimized output Q^* is larger than available capacity. In these cases, we need a rationing rule, for which we use a "pro-rata" rule. The pro-rata rule reflects possible asymmetry ($\gamma_1 \neq \gamma_2$), which would be neglected with an "equal-split" rule.

Define

$$\omega_{B1} = \frac{Q_{B1}^*}{Q^*} \text{ and } \omega_{B2} = \frac{Q_{B2}^*}{Q^*} \tag{16}$$

[7] Note however, that even in the high capacity case we use the "low" capacity for B_2 because we assume that B_2's capacity is fixed at the low level, which implicitly relies on $\gamma_1 > 1$.

Denote "**" as the contrained optimized (post-rationing) outcome. Then:

$$Q^{**}_{B1} = Q^{*}_{B1} - \omega_{B1}(Q^{*} - K_A) \text{ and } Q^{**}_{B2} = Q^{*}_{B2} - \omega_{B2}(Q^{*} - K_A) \qquad (17)$$

Note that the rationing rule is only relevant for the low network-capacity case (K_A^L). In the high network-capacity case (K_A^H), capacity constraints cannot occur by construction.

Using α to weigh the different states of demand, we can now construct a payoff table, as in table **1**, with which we can construct and solve the capacity choice problem. Table **1** gives the profits for A (down left) and generator B_1 (top right), for constrained optimized values of Q and capacity choices low and high.

		B_1	
		B_1 low	B_1high
A	low	Π_A^{LL} / Π_{B1}^{LL}	Π_A^{LH} / Π_{B1}^{LH}
	high	Π_A^{HL} / Π_{B1}^{HL}	Π_A^{HH} / Π_{B1}^{HH}

Table 1: General structure of the payoff matrix for A and B_1

Recall our assumption that B_2 does not choose capacity; therefore, B_2 does not show up as a strategic player in the table (but of course, the pay-offs do reflect optimization of Q_{B2}). The table is constructed as a reduced form game-theoretical model, where, to recall, B_1 signals L or H first, after which A chooses K_A^L or K_A^H and then B_1 chooses K_{B1}^L or K_{B1}^H.

Incentive compatibility

The last step is to specify the "cheap-talk" problem, or in other words, the conditions for a situation where costless information exchange is not credible. We investigate whether the investment problem can be solved by exchanging information: A can simply ask B_1 what it will do and B_1 responds acoordingly. However, if B_1 may lie, the information may not be credible and therefore the information exchange may be ineffective if A does not know whether he can believe B_1 or not. In more formal terms, cheap talk as a device to resolve the information problem may not be incentive compatible.

We refer to table 1 to specifiy the cases. The relevant case happens when B_1 prefers a large network K_A^H, while B_1 wants to invest in low capacity K_{B1}^L itself (given a large network), while at the same time, the network operator would only invest in a large network if B_1's capacity is high, and would invest in small network if B_1's capacity is low. In this case, B_1 has an incentive to lie. It would signal that it will invest in high capacity, $\sigma^H(K_{B1})$, to trigger a large network, but would actually invest in low capacity, K_{B1}^L, after investment in a large network has been made. This situation, where cheap talk is not credible and thus incentive compatibility is violated, translates into two conditions:

Condition 1: No dominant strategy for the network.

This requires that neither $\left\{\pi_A^{LL} \geq \pi_A^{HL}, \text{ and}, \pi_A^{LH} \geq \pi_A^{HH}\right\}$, nor, $\{\pi_A^{LL} < \pi_A^{HL}, \text{ and}, \pi_A^{LH} < \pi_A^{HH}\}$ exists.

Formally, for condition 1 we require:

$$\left\{\pi_A^{HH} > \pi_A^{LH}, \text{ and}, \pi_A^{LL} > \pi_A^{HL}\right\} \tag{18}$$

Note that the alternative constellation of non-dominance for the network: $\left\{\pi_A^{LL} < \pi_A^{HL}, \text{ and}, \pi_A^{HH} < \pi_A^{LH}\right\}$ does not exist.[8]

Condition 1 is not directly related to incentive compatibility, but secures that the problem is non-trivial. If the network owner would face a dominant strategy, the problem would be gone and information exchange would be useless and meaningless.

Condition 2: Violation of incentive compatibility for B_1, (given that condition 1 is fulfilled)

This condition requires that:

$$\left\{\pi_{B1}^{HL} > \pi_{B1}^{LL}, \text{ and}, \pi_{B1}^{HL} > \pi_{B1}^{HH}\right\} \tag{19}$$

If condition 2 is fulfilled, B_1 will signal high capacity $\sigma^H(K_{B1})$, enforcing a large network (if A believes the signal), K_A^H, and then, given the large network, B_1 will actually invest in low capacity K_{B1}^L. Note that B_1 must lie to get to this result. If it signals low capacity, A will invest in a small network (due to condition 1).

The two legs of condition 2 are equivalent to definition 3 in Baliga and Morris (2002, p.455), but then formulated as a violation rather than a confirmation of incentive compatibility. For cases where conditions 1 and 2 are fulfilled cheap talk does not work, and simple information exchange does not adequately address the investment problem. We show in the next section, that these cases, where lying is profitable, do exist, and that an integrated solution would be different and in fact welfare improving.

We compare the vertically separated (unbundled) case with the vertically integrated case. For the vertically integrated approach we take joint-profit maximization, for which we simply use the sum of the profits of the separate parts. We do not separately optimize for the integrated solution. The reason is that using the sum of separate parts allows for better comparison of pure separation effects. In a new, separately optimized joint-profit solution, we lose competition (among B_1 and B_2), in which case it is no longer clear what exactly is being

[8] $\pi_A^{LL} < \pi_A^{HL}$ would require $2\,(1-\gamma_1)\,c_B > (a^H - a^L)$, which is never true as $\gamma_1 \geq 1$ and $a^H \geq a^L$. Therefore we can conclude that the alternative condition for non-dominance does not exist.

compared. Vertical separation may cause cost of coordination which may be offset by improved competition, and therefore we would lose information on the coordination problem, which is the focus of this paper.

For social welfare we follow convention and calculate the unweighted sum of consumer and producer surplus. We do not maximize social welfare; we only compare the profit-driven solutions under vertical separation and vertical integration and compare these outcomes on social welfare.

Asymmetry and symmetry compared

We have made the explicit assumption of asymmetry in the effects of network expansion: one firm can gain a competitive advantage over its competitor from network expansion by the cost-lowering change in γ_1. The asymmetry is the positive spillover, which drives the model. This begs the question what happens under symmetry. In the setting of our model this is a non-trivial exercise. We would expect that in a perfectly closed model with perfect information and rationality, perfect cost-incidence would secure that the spillover effects vanish. In other words, if everyone knows that the network expansion costs will be passed on to the end-users and will thereby affect final demand, and assuming that all symmetric players know and expect this and act accordingly, all costs and benefits are shared symmetrically and it is no longer obvious why it would be profitable to set a false signal. For our context, there are two problems with this line of argument. First, what should be assumed about the reactions of players to the symmetric cost-lowering effect of network expansion? If we assume that all players know that this effect applies to all players, we should model that all players recalculate optimal strategies of the competitors, which contradicts the Cournot assumption. Alternatively, we may assume and model that players think they can get a competitive advantage, and thus keeping quantities of others constant. This is actually realistic, but leads again to some kind of asymmetry. Second, our model is not entirely closed but actually has a leakage. By the assumption about the used-and-useful regulation of the network expansion, the risk that the expansion is not used rests with the network owner. If the expansion is not used, it is a loss for shareholders. This means in other words that in our approach there is no perfect cost-incidence. Both these aspects imply that we would have to change the setting significantly to model symmetry and comparison with the case of asymmetry would no longer possible. Therefore, we restrict attention to asymmetry.

4 Results

We are now ready to present and discuss the main results. Although the main ideas are actually intuitive, we formulated the central conclusions in formally proven propositions. Several propositions claim that "a situation exists", implying that a numerical example showing existence of this situation is sufficient as a proof.

Proposition 1 *Strategic investment withholding. Provided incentive compatibility is given, the case of full separation can enforce a welfare improving state as compared to the (joint profit) integrated case. This happens if players that choose capacity, including network size, gain less from capacity expansion, than other players lose. The underlying rationale is that integration is subject to incentives of strategic investment withholding.*

Proof. We have explained above in section 2 that a key argument in the network unbundling debate is strategic investment withholding, which is perfectly illustrated with the numerical example underlying proposition 1. Strategic investment withholding basically states that a vertically integrated company may have an incentive not to expand network capacity so as to protect its competitive business from competitors. Table 2 denotes the parameter values we use for proposition 1. Table 3 gives profits for the network A and generator B_1.

Parameters			
a^L	300	n	2
a^H	320	c_A	12
b	1	β_A	30
α	$0,8$	c_B	5
z_1	0	β_B	5
z_2	0	γ_1	$1,4$
μ	63	γ_2	$1,4$

Table 2: Parameter values used for proposition 1

		B_1	
		B_1 low	B_1 high
A	low	6721 / 3220	6928 / 3220
	high	6874 / 2980	6935 / 3313

Table 3: Profits for A and B_1

Clearly, the separated outcome is high-high. B_1 will signal high capacity, after which A will expand the network, and it is the self-interest of B_1 to invest in high capacity. This is an ideal outcome where information exchange does work. Table 4 gives profits for A and B_2. Recall that B_2's capacity is always low, as by assumption B_2 does not choose its capacity. Therefore, the table 4 only depicts the profits for B_2 under different scenarios.

Table 5 gives joint profits in case of an integrated solution in the upper right corner and social welfare (sum of surpluses) in the bottom left of each cell. From table 5 we see that an integrated firm would opt for the low-low case as this maximises joint profits. We also see from table 5 that the low-low case gives

		B_2	
		B_1 low	B_1high
A	low	6721 / 3220	6474 / 3220
	high	6721 / 2980	6210 / 3313

Table 4: Profits for A and B_2

		Joint profits	
		B_1 low	B_1high
Social Welfare	A low	16662 / 28418	16622 / 28378
	A high	16576 / 28331	16458 / 29241

Table 5: Producer surplus (joint profits of A, B_1 and B_2; top right) and social welfare (bottom left)

lower social welfare than the high-high case, for these parameter values. Thus we conclude, that players A and B_1 in the case of separation coordinate on a high-high outcome (in their self-interest), which is different from and welfare-improving as compared to the integrated outcome. The reason is that the gains of capacity expansion (high-high as compared to low-low) for A and B_1 is less than the associated loss for B_2. In an integrated setting, B_2 would be part of A, and therefore, the firm foregoes additonal profits at A and B_1 to avoid the losses at B_2. It would thus not invest in network expansion in order to avoid stronger competition from B_1 on B_2. In a world with unbundling, A and B_1 would not consider the profits effects on B_2. This is precisely what the argument of strategic investment withholding states. **End of proof**

Whereas proposition 1 gives a fundamental result of the unbundling debate, but is not per se related to the coordination or information problem, we now turn to the following propositions to explore just this. First, table 6 gives the used parameters; note that these are different from the parameter values for proposition 1.

Proposition 2 *There exists a situation where cheap talk is not incentive compatible.*

Proof Table 7 gives the profits for players A and B_1 depending on capacity choices. First note that the table fulfills condition 1, because the network investor does not have a dominant strategy; its optimal choice depends on B_1. B_1 would maximize its profits in the cell bottom-left (i.e a large network, but

Parameters			
a^L	300	n	2
a^H	320	c_A	12
b	1	β_A	30
α	$0,6$	c_B	5
z_1	0	β_B	30
z_2	0	γ_1	$1,4$
μ	63	γ_2	$1,4$

Table 6: Parameter values used for propositions 2 - 4

low generation capacity). However, if it reveals this preference and signals low generation capacity, A will invest in a small network and the game ends up in the cell top-left (low-low). The only strategy for B_1 is to lie: it will signal high generation capacity, triggering A to invest in network expansion, and then B_1 will not invest in high generation capacity. Therefore, under these parameters cheap talk is not incentive compatible. **End of proof.**

		B_1	
		B_1 low	B_1high
A	low	3220 \ 4498	3220 \ 4443
	high	2980 \ 4651	3238 \ 4499

Table 7: Profits for A and B_1

Proposition 2 is the main claim of this paper. It states that we cannot generally rely on straightforward information exchange to solve the investment coordination problem that is created by unbundling. The fact alone that cheap talk may not be incentive compatible means that we can never be sure whether information exchange works or not, because A would never know whether B_1 lies or tells the truth. This is essentially proposition 10 in Baliga and Morris (2002, p.462)[9]. Therefore, we need an external coordination device to address the problem adequately.

Proposition 3 *Given the violation of incentive compatibility in proposition 2, there exists a situation where the (profit-optimized) non-integrated outcome differs from the (profit-optimized) integrated outcome.*

[9]It may be noted that the structure of the problem above mirrors the structure of the high-cost situation in figure 6 in Baliga and Morris (2002, p.458), where "truth-telling is no longer an equilibrium." As the structure is the same, the claim should be the same as well and therefore, proposition 10 in Baliga and Morris (2002, p.462) applies, which mirrors our proposition 2.

Proof: With the parameters from table 6, resulting in missing incentive compatibility as shown in table 7, table 8 presents joint profits (top right) and social welfare (bottom left) for this situation. It is clear from table 8 that the fully integrated firm would opt for the low-low outcome, as this maximizes the sum of profits. Therefore, if we say that under these parameters and if A would believe the signal of B_1, the outcome under vertical separation will be high network capacity and low-generation capacity (see table 7), we have now established the result, that the separated outcome will be different from the integrated outcome. **End of proof**.

This proposition is not central to our main claim, but we need to show that there is actually a difference between separation and integration. It could have been the case, that under separation B_1 lies to get to an outcome which is exactly the same as under integration; in that case, lying would actualy be good, and we would say that lying apparently repairs the coordination problem. Proposition 3 shows that this is not generally so, because there is at least one case where the separated and integrated outcome differ.

		Joint profits	
		B_1 low	B_1high
Social Welfare	A low	12216 (top right) / 23971 (bottom left)	11976 (top right) / 23731 (bottom left)
	A high	12129 (top right) / 23884 (bottom left)	11827 (top right) / 24379 (bottom left)

Table 8: Producer surplus (joint profits of A, B_1 and B_2; top right) and social welfare (bottom left)

Proposition 4 *Given propositions 2 and 3, there exists a situation where the integrated situation is welfare-improving (in social surplus) as compared to the non-integrated situation.*

Proof: Proposition 2 claims that, provided that A believes B_1's signal, the separated outcome will be bottom-left (i.e high network capacity and low generation capacity). Proposition 3 claims that the integrated outcome will result in the top-left outcome (low network capacity and low generation capacity), indicating that A will not expand the network if it would know that generation capacity will be low. Applying table 8, we immediately see that the top-left is welfare-improving as compared to bottom-left. Thus we conclude for these parameters that separation decreases social welfare. **End of proof.**

This proposition is crucial. It implies that the outcome achieved with lying can actually be bad for social welfare. This is important, because in principle, even if we have an uncoordinated, not incentive-compatible outcome (prop. 2), and even if this differs from the integrated solution (prop. 3), the outcome might still be better for social welfare than the integrated case. Proposition 4 shows that this is not generally the case and that there is at least one case where separation would decrease social welfare. In this case, we thus have a genuine case of costly coordination due to vertical separation.

We should stress though that proposition 4 does not make a general statement about the pros and cons of network unbundling. It merely states that unbundling can cause coordination costs. We have only shown that problems may arise in case of positive spill overs. Moreover, we have explicitly not modelled the positive effects on competition. Therefore, proposition 4 does not make a statement on what happens on balance.

Summing the three propositions 2, 3, and 4 above, we have to conclude that network unbundling can indeed cause an investment coordination problem, that is not easily resolved by simple exchange of information, and that is welfare decreasing. Put differently, there are costs of coordination due to a suboptimal outcome. In effect, we conclude that where network unbundling removes firm-internal coordination, we need explicit market design to implement external coordination. More specifically, network charging can signal and steer investment needs. Network charging is a matter of regulation, which we explore as a coordinating device in the following propositions.

The problem above is essentially that B_1 may benefit from a costly network expansion without paying for it. In other words, this is a free-riding problem. If this is the problem, then the straightforward answer appears to be cost-reflective pricing, meaning that the beneficiary of the network expansion should pay for it. In electricity markets this is generally called "deep charging". The alternative is called "shallow charging", where a new connection (say, a power plant) to the network only pays the costs of actually connecting to the network but not for network upgrades beyond the point of connection. If the new connection also pays for the cost of a network upgrade beyond the point of connection (deeper in the network), then this is called deep charging. Deep charging can be fully cost reflective, when all costs are attributed to the causer, or partial if only part of these costs are attributed to the beneficiary. If all costs are passed through to the beneficiary we call this full-cost deep charging.

In our approach above, the deep charge is z_1 for B_1, and simultaneously, by definition, the deep charge is a deep compensation z_2 for B_2, in order to establish revenue neutrality for the network A. We can see the effect of the deep charge by examining the profit expressions for π_{B1}^{LL} and π_{B1}^{HL}(detailed in the appendix). If z_1 is increased (stronger deep charging), π_{B1}^{HL} goes down, whereas π_{B1}^{LL} is not affected (because under LL the network is not expanded). Therefore, increasing z_1, implies that the first part of condition 2 is less likely to be fulfilled; in other words, increasing z_1 makes it more likely that cheap talk is incentive compatible. This is intuitive, because it is less attractive to free-ride if you have to pay for the ride.

The more interesting question is how high the deep charge z_1 should be to solve the incentive problem. At first glance, we would expect that a full-cost deep charge would exactly solve the problem; unfortunately this is not so. We address this point in proposition 5.

Proposition 5 *The minimum charge that internalizes the incentive compatibility problem is either lower or higher than full-cost deep charging. In other words, even full-cost deep charging may not adequately address the incentive compatibility problem.*

Proof: In total we have to show the conditions 1 and 2 specified above.

First, we specifiy condition 1:

1.1. $\pi_A^{HH} > \pi_A^{LH}$
1.2. $\pi_A^{LL} > \pi_A^{HL}$

Using the profit equations in the appendix, equalizing and working out then gives for condition 1.1:[10]

$$(\mu - c_A)\left[\alpha\left(a^H - a^L + (\gamma - 1)\, c_B\right) + (\gamma - 1)\, c_B\right] > \beta_A\left(a^H - a^L + 2(\gamma - 1)c_B\right) \tag{20}$$

For α close to 1 this means that capacity cost have to be smaller than $\mu - c_A$. As by definition $\mu = \varepsilon\,(c_A + \beta_A)$, with ε higher than 1, this is true, if ε and/ or α is sufficiently high. The possible extended output realized with probability α and the output increase realized because of B's output increase from reduced cost under high network capacity have to outweigh the additional cost for building the high capacity.

The requirement for condition 1.2 is:

$$\left(a^H - a^L\right) > 2(1 - \gamma)c_B \tag{21}$$

which is always fulfilled as $\gamma \geq 1$ and $a^H > a^L$.

Second, and more importantly, we specify condition 2 (as formulated here, violates incentive compatibility):

2.1 $\pi_{B1}^{HL} > \pi_{B1}^{LL}$
2.2 $\pi_{B1}^{HL} > \pi_{B1}^{HH}$

Condition 2.2 specifies the cases in which own capacity investment does not pay off for B_1. This occurs if capacity costs are bigger then possible profit increases from expanded output corrected for effects from price changes.[11]

[10] As γ is either $\gamma_1 = \gamma_2 > 1$ or $\gamma_1 = 1$ and $\gamma_2 > 1$ we have simplified expressions dropping γ_1 whenever $\gamma_1 = 1$ and writing γ whenever $\gamma_1, \gamma_2 > 1$ appears.

[11] Formally the condition implies $\beta_B\left(a^H - a^L + 2(\gamma_1 - 1)c_B\right) > \frac{\alpha}{3}\left[\left(a^H - a^L\right)\left[2\left(a^H - a^L\right) + (5\gamma_1 - 2)\, c_B\right] - \left(a^L - (2 - \gamma_2)\, c_B - \mu\right)\gamma_1 c_B\right] + \frac{1}{3} c_B\,(\gamma_1 - 1)\left(a^L - (4 - 2\gamma_1 - \gamma_2)\, c_B - \mu\right)$.
The probability of high demand plays an important role as only than capacity expansion does translate in equivalent output. In case of low demand only a marginal increase over low capacity can be realized as a result of the positive spillover.

For condition 2.1, we want to find z_1^*, which we define as the value of z_1 which exactly repairs incentive compatibility. In other words, z_1^* equalizes both sides of condition 2.1. Using the expressions in the appendix and solving for z_1^*, then gives:

$$z_1^* = \frac{a^L - \gamma c_B - \mu}{3b}(\gamma - 1)c_B \tag{22}$$

To solve the problem of incentive compatibility, z_1 has to be sufficiently high, i.e. $z_1 > (\gamma - 1)c_B \frac{\left(a^L - \gamma c_B - \mu\right)}{3b}$, which is the same as $z_1 > (\gamma - 1)c_B K_{B1}^L$. Define the expansion costs as $\left(K_A^H - K_A^L\right)\beta_A$. We now compare the incentive compatible charge z_1^* to the expansion costs. Since $K_A^L = \frac{2\left(a^L - \gamma c_B - \mu\right)}{3b}$ and $K_A^H = \frac{\left(a^H + a^L - 2c_B - 2\mu\right)}{3b}$, we find that:

$$\left(K_A^H - K_A^L\right)\beta_A = \beta_A \frac{a^H - a^L + 2(\gamma - 1)c_B}{3b} \tag{23}$$

Then $z_1^* < \left(K_A^H - K_A^L\right)\beta_A$, implies

$$(\gamma - 1)c_B \frac{\left(a^L - \gamma c_B - \mu\right)}{3b} < \beta_A \frac{a^H - a^L + 2(\gamma - 1)c_B}{3b} \tag{24}$$

which solves to:

$$\frac{\left(a^L - \gamma c_B - \varepsilon c_A\right)(\gamma - 1)c_B}{\left(\varepsilon(\gamma - 1)c_B + \left(a^H - a^L + 2(\gamma - 1)c_B\right)\right)} < \beta_A, \text{with } \mu = \varepsilon\left(c_A + \beta_A\right) \tag{25}$$

This inequality is not unambiguously true as can readily be seen. All the terms in the LHS are unambiguously positive, with the exception of the first top-left term in round brackets, which can be negative and positive. If we assume $\beta_A = 0$, it is then clear that the inequality is not unambiguously true. Therefore, proposition 5 holds: the deep charge that restores incentive compatibility can be higher or lower than full-cost deep charging.**End of proof.**

The result of proposition 5 is important. The structure of the cheap-talk problem does not perfectly coincide with the effects on the cost of network expansion. Upon reflection this is intuitive. The ultimate reason for the incentive-compatibility problem is the positive spillover, and not the network expansion costs. The effect of the spillover depends on the (asymmetric) cost advantage (the change in γ_1) and competitive conditions. A closer look at the conditions derived above suggests the following. If β_A is low, z^* is likely to be larger than network expansion costs and thus the deep charge would be higher than the network expansion costs. At the same time, if β_A is low, the coordination problem stops being relevant, simply because a too large network would not be costly and therefore the distortive effect would be small. If β_A is high, the condition above is more likely to be fulfilled and it is more likely that z^* is smaller than network expansion costs. This appears to be the more realistic case. Therefore, we conclude that normally the minimal deep charge is lower than network expansion costs, or, where this is not the case, there is no relevant problem and in that case it might be best not to use a deep charge in the first place. The approach above seems to suggest that overall a "deep-ish" charge might work perfectly well and mitigate strategic signalling.

5 Dissussion and concluding remarks

This paper provides a formal analysis on the investment coordination problem in a vertically unbundled electricity supply industry. Thereby the paper specifies the argument that vertical unbundling causes "coordination costs" or "transaction costs". Quite often this argument is either simply taken for granted and stays on a rather general level, without specifying what these coordination costs exactly are. While we focus on the high-voltage transmission grid in the electricity sector, we note that the analysis may also apply to low-voltage distribution grids, and to other network industries. Also, this paper does not make a conclusive statement on the balance of costs and benefits of (ownership) unbundling. We focus on some aspects of the costs of coordination and do not analyse the competition effect of unbundling.

The problem we examine is the following. In an electricity system, network development depends on the location and capacity of the power plants connected to the network and the other way around. In other words, to optimize the system, the investment decisions of network and generators need to be coordinated. In unbundled, or more generally liberalized markets, the investment decisions are decentralized. Therefore, firm-internal coordination no longer appplies and should be replaced by external market coordination. The coordination problem is twofold: an information problem and an incentive problem. If the problem is that investment plans of the generators are unknown to the network planner, the straightforward answer would be that the network planner would simply ask the generators. The obvious counterargument is that the generators might have an incentive to lie. We develop a formal approach to examine whether simple information exchange could address the investment coordination problem. In other words, can "cheap talk" solve the investment coordination problem?

To address this question we adopt a closed three-stage profit optimized investment model, with a (regulated) monopoly network and two asymmetrical Cournot-type generators. In stage 1, the generators signal whether they invest in high or low generation capacity. In stage 2, the network planner uses this signal to decide whether to invest in large or small network capacity. In stage 3, having seen the network investment decision, the generators decide whether to invest in large or small generation capacity. Importantly, the generators may lie. The model then formally examines whether the generators have an incentive to lie at all.

Formally, this leads to a game-theoretical analysis of the credibility of cheap talk. We follow Aumann (1990) and Baliga and Morris (2002) for "self-signalling" as our lead criterion to evaluate credibility of cheap talk. Reflecting the practical problem we address, our approach is a cheap-talk game with one-way communication, incomplete information, and positive spillovers. The structure of the theoretical approach in Baliga and Morris (2002) mirrors ours nicely such that we can adopt their definitions, propositions and proofs. In particular, we use definition 3 in Baliga and Morris (2002, p.455), which sets out the conditions to be fulfilled for incentive compatibilty.

First, the model nicely shows the existence of what is called "strategic investment withholding", which is one of the key arguments in the unbundling debate. The argument claims that vertically integrated firms may have insufficient incentives to invest in network capacity as this could strengthen competition for their commercial parts. A fully unbundled network owner would not have to defend financial interest in these commercial businesses and would thus have different incentives for network investment. With proposition 1, the model gives an example where full separation can enforce a welfare improving state as compared to the (joint profit) integrated case. This happens if players that choose capacity, including network size, gain less from capacity expansion, than other players lose.

Second, and this is the core of the paper, we show in proposition 2 that cases exist that violate the self-signalling condition of cheap-talk credibility; therefore, unfortunately, cheap talk as a coordination device breaks away. Propositions 3 and 4 follow up on the same case of proposition 2 and show that an integrated (profit maximizing) firm would act differently and that the integrated situation would be welfare improving compared to the case of separation. With this we have shown that cheap talk cannot generally solve the investment coordination problem and that as a result separation may actually cause a costly coordination problem.

We stress that these conclusions do not make a conclusive statement on the balance of the costs and benefits of ownership unbundling. The mere point is that there may be an investment coordination problem. We did not include the competition benefits of unbundling in our analysis and therefore we cannot draw conclusions with regard to the overall effect.

The last part of the paper steps into network pricing as a coordination device. We examine what is known as "deep charging". Underlying the incentive-compatibility problem of cheap talk is what Baliga and Morris (2002) call a positive spillover. Our positive spillover is that network expansion benefits the generators. Therefore the generators have an incentive to signal large investment plans to trigger network expansion, even if they do not actually invest in large generation capacity. If this is the problem, then the obvious solution is to make generators pay for the network expansion on their behalf. Cost-reflective charging for network reinforcement to facilitate new generator connections is called "deep charging". With proposition 5, we show that full cost-reflective deep charging, most unfortunately, does not repair the problem that cheap talk fails. The problem of failing cheap talk is the spillover, not the cost of network reinforcement. More precisely, there is a range where the deep-ish charge which would repair incentive compatibility of cheap talk is below full cost-reflective deep charging. This is good news, as full cost-reflective deep charging is problematic in practice.[12] This situation where the incentive-compatible deep charge is lower than full cost underlines the idea that the spillover is the problem and not the network reinforcement cost. Instead of deep charging, one could consider a "down-payment" for generator investment signals as a self-commitment device; if a generator signals high capacity which requires network reinforcement the network owner could ask for a down-payment. If later, the generator steps back on its decision and actually invested in low capacity, the down-payment would be lost. A well-chosen down-payment could internalize the incentive problem and could thus serve as an investment coordination device.

[12]Unfortunately though the reverse may also hold: the deep-ish charge which would repair incentive compatibility of cheap talk is above full cost-reflective deep charging. However, this case does not seem to be very relevant.

6 Appendix

This appendix lists the analytical solutions for the profits in table 1.

First, analytical expressions for the solutions of capacities are:

$$
\begin{aligned}
K_{B1}^{H} &= \frac{\left(a^{H} - (2-\gamma_2)c_B - \mu\right)}{3b} \\
K_{B1}^{L} &= \frac{\left(a^{L} - (2\gamma_1-\gamma_2)c_B - \mu\right)}{3b} \\
K_{B2} &= \frac{\left(a^{L} - (2\gamma_2-\gamma_1)c_B - \mu\right)}{3b} \\
K_{A}^{L} &= \frac{\left(2a^{L} - (\gamma_1+\gamma_2)c_B - 2\mu\right)}{3b} \\
K_{A}^{H} &= \frac{\left(a^{H} + a^{L} - (2-\gamma_1+\gamma_2)c_B - 2\mu\right)}{3b}
\end{aligned}
$$

Note that we have left the expressions γ_1 and γ_2 to avoid confusion, although it can be further simplified. Either $\gamma_1 = 1$ which has been substituted in these expressions or $\gamma_1 = \gamma_2 > 1$, which has not been (but can be) substituted. Therefore, where γ_1 shows up in the expressions below, it necessarily means that $\gamma_1 = \gamma_2 > 1$.

Note that γ_1 still shows up in the case of high network capacity. This is because $K_A^H = Q_A^{H*} = Q_{B1}^{H*} + Q_{B2}^{L*}$, where Q_{B2}^{L*} relies on $\gamma_1 > 1$.

Below we present the calculated analytical expressions for profits. To get the final outcome, low and high demand outcome have been summed up weighted with α for the high demand solution and $(1-\alpha)$ for low demand. In the background calculations, sometime for technical reasons we have to distinguish different cases depending on whether $(\gamma_1 - 1)\,c_B$ is larger or smaller than $\left(a^H - a^L\right)$. For the expression below, we use $(\gamma_1 - 1)\,c_B \leq \left(a^H - a^L\right)$ as the more likely condition for a wide set parameter values of γ, c_B, a^H and a^L; the other options of the case differentiation have been dropped for simplicity. The resulting solutions for the respective profits are listed below:

Low network and low generation investment, π^{LL}

$$\pi_A^{LL} = \frac{(\mu - c_A - \beta_A)(2a^L - (\gamma_1+\gamma_2)c_B - 2\mu)}{3b}$$

$$\pi_{B1}^{LL} = \left(\alpha(a^H - a^L) + \frac{(a^L - (2\gamma_1-\gamma_2)c_B - \mu)}{3} - \beta_B\right)\frac{a^L - (2\gamma_1-\gamma_2)c_B - \mu}{3b}$$

$$\pi_{B2}^{LL} = \left(\alpha(a^H - a^L) + \frac{(a^L - (2\gamma_2-\gamma_1)c_B - \mu)}{3} - \beta_B\right)\frac{(a^L - (2\gamma_2-\gamma_1)c_B - \mu)}{3b}$$

Low network and high generation investment, π^{LH}

$$\pi_A^{LH} = \frac{(\mu - c_A - \beta_A)(2a^L - (\gamma_1 + \gamma_2)c_B - 2\mu)}{3b}$$

$$\pi_{B1}^{LH} = \left(\alpha(a^H - a^L) + \frac{(a^L - (2\gamma_1 - \gamma_2)c_B - \mu)}{3}\right) \cdot \frac{a^L - (2\gamma_1 - \gamma_2)c_B - \mu}{3b} - \beta_B\left(\alpha\frac{a^L - a^H - (2\gamma_1 - 2)c_B)}{3b} + \frac{(a^H - (2 - \gamma_2)c_B - \mu)}{3b}\right)$$

$$\pi_{B2}^{LH} = \alpha\frac{(3a^H - 2a^L + (\gamma_1 - 2\gamma_2)c_B - \mu)}{3b}\frac{(2a^L - (\gamma_1 + \gamma_2)c_B - 2\mu)}{(a^H + a^L - (\gamma_1 + \gamma_2)c_B - 2\mu)} \cdot \frac{(a^L - (2\gamma_2 - \gamma_1)c_B - \mu)}{3b} - \alpha\frac{(a^L - (2\gamma_2 - \gamma_1)c_B - \mu)}{3}\frac{(a^L - (2\gamma_2 - \gamma_1)c_B - \mu)}{3b} + \left(\frac{(a^L - (2\gamma_2 - \gamma_1)c_B - \mu)}{3} - \beta_B\right)\frac{(a^L - (2\gamma_2 - \gamma_1)c_B - \mu)}{3b}$$

High network and low generation investment, π^{HL}

$$\pi_A^{HL} = \frac{(\mu - c_A)\left(2a^L - (\gamma_1 + \gamma_2)c_B - 2\mu\right)}{3b} - \beta_A\frac{(a^H + a^L - (1 + \gamma_2)c_B - 2\mu)}{3b}$$

$$\pi_{B1}^{HL} = \left(\alpha(a^H - a^L) + \frac{(a^L + (\gamma_1 + \gamma_2 - 3)c_B - \mu)}{3} - \beta_B\right) \cdot \frac{a^L - (2\gamma_1 - \gamma_2)c_B - \mu}{3b} - z_1$$

$$\pi_{B2}^{HL} = \left(\alpha\left(a^H - a^L\right) + \frac{\left(a^L - (2\gamma_2 - \gamma_1)c_B - \mu\right)}{3} - \beta_B\right) \cdot \frac{a^L - (2\gamma_2 - \gamma_1)c_B - \mu}{3b} - z_2$$

High network and high generation investment, π^{HH}

$$\pi_A^{HH} = \alpha \frac{(\mu - c_A)(a^H - a^L - (1-\gamma_1)c_B)}{3b} + \frac{(\mu - c_A)(2a^L - (1+\gamma_2)c_B - 2\mu)}{3b} - \beta_A \frac{(a^H + a^L - (2-\gamma_1+\gamma_2)c_B - 2\mu)}{3b}$$

$$\pi_{B1}^{HH} = \alpha \left(\frac{(2a^H - a^L - (\gamma_1 + 2 - \gamma_2)c_B - \mu)}{3} \right) \frac{(a^H - (2-\gamma_2)c_B - \mu)}{3b} + (1-\alpha) \frac{(a^L - (2-\gamma_2)c_B - \mu)^2}{9b} - \beta_B \frac{(a^H - (2-\gamma_2)c_B - \mu)}{3b} - z_1$$

$$\pi_{B2}^{HH} = \left(\alpha \frac{(2a^H - a^L - (\gamma_1 - 2 + 2\gamma_2)c_B - \mu)}{3} - \beta_B \right) \cdot \frac{(a^L - (2\gamma_2 - \gamma_1)c_B - \mu)}{3b} + (1-\alpha) \frac{(a^L - (2\gamma_2 - 1)c_B - \mu)^2}{9b} - z_2$$

References

Ackermann, T., 2004. Distributed Resources in a Re-Regulated Market Environment. Ph.D. thesis. KTH, Stockholm.

Aumann, R.J., 1990. Nash equilibria are not self-enforcing. Economic decision-making: Games, Econometrics and Optimisation: Contributions in Honour of Jacques H. Dreze , 201–06.

Aumann, R.J., Hart, S., 2003. Long cheap talk. Econometrica 71, 1619–1660.

Baldick, R., Kahn, E., 1993. Network costs and the regulation of wholesale competition in electric power. Journal of Regulatory Economics 5, 367–384.

Baliga, S., Morris, S., 2002. Co-ordination, spillovers, and cheap talk. Journal of Economic Theory 105, 450–468.

Balmert, D., Brunekreeft, G., 2010. Deep ISOs and network investment. Competition and Regulation in Network Industries 1, 27–49.

Brunekreeft, G., McDaniel, T.M., 2007. New Energy Paradigm. Oxford University Press. Policy Uncertainty and Supply Adequacy in Electric Power Markets, Ch. 11. pp. 320–347. Updated version of article with same title in Oxford Review of Economic Policy, 2005, Vol. 21, No. 1, pp. 111-127.

Brunekreeft, G., Neuhoff, K., Newbery, D., 2005. Electricity transmission: An overview of the current debate. Utilities Policy 13, 73–93.

Crawford, V.P., Sobel, J., 1982. Strategic information transmission. Econometrica 50, 1431–1451.

DENA, 2005. Energiewirtschaftliche Planung für die Netzintegration von Windenergie in Deutschland an Land und Offshore bis zum Jahr 2020. Deutsche Energie Agentur (DENA).

ETSO, 2003. Report on renewable energy sources (RES). European Transmission System Operators.

European Commission, 2007. Report on energy sector inquiry SEC(2006)1724, 10 January 2007. European Commission, DG Competition.

Farell, J., 1988. Communication, coordination and nash equilibrium. Economics Letters 27, 209–214.

Farell, J., 1993. Meaning and credibility in cheap-talk games. Games and Economic Behavior 5, 514–531.

Farell, J., Rabin, M., 1996. Cheap talk. The Journal of Economic Perspectives 10, 103–118.

Hogan, W.W., 2008. Electricity Market Structure and Infrastructure, in: Conference on Acting in Time on Energy Policy, Harvard University, September 18-19, 2008, Springer. pp. 211–242.

IEA, 2007. Tackling investment challenges in power generation in iea countries: Energy market experience. International Energy Agency, Organization for Economic Co-operation and Development (OECD).

Joskow, P., Schmalensee, R., 1983. Markets for Power: An Analysis of Electric Utility Deregulation. MIT Press.

Kwoka, J.E., 2002. Vertical economies in electric power: Evidence on integration and its alternatives. International Journal of Industrial Organization 20, 653–671.

Li, F., 2007. The benefit of a long-run incremental pricing methodology to future network development, in: IEEE Power Engineering Society General Meeting, 24-28 June 2007, Tampa, Florida USA.

Li, F., Tolley, D., Padhy, N.P., Wang, J., 2005. Network Benefits from Introducing an Economic Methodology for Distribution Charging. Technical Report. Department of Electronic and Electrical Engineering, University of Bath.

Nemoto, J., Goto, M., 2004. Technological externalities and economies of vertical integration in the electric utility industry. International Journal of Industrial Organization 22, 67–81.

OFGEM, 2009. Electricity distribution structure of charges project: The common distribution charging methodology at lower voltages. Decision Document 140/09. Office of the Gas and Electricity Markets.

Prica, M., Ilic, M.D., 2007. Optimal distribution service pricing for investment planning, in: IEEE Power Engineering Society General Meeting, 2007, Citeseer. pp. 1–7.

Scheepers, M., Bauknecht, D., Jansen, J., de Joode, J., Gómez, T., Pudjianto, D., Ropenus, S., Strbac, G., 2007. Regulatory improvements for effective integration of distributed generation into electricity distribution networks – Summary of the DG-GRID project results.

Schweppe, F.C., Tabors, R.D., Caramanis, M.C., Bohn, R.E., 1988. Spot Pricing of Electricity. Kluwer Academic Publishers, Norwell, MA.

Strbac, G., Mutale, J., 2007. Framework and Methodology for Pricing of Distribution Networks with Distributed Generation - A report to OFGEM. Technical Report. Centre for Distributed Generation and Sustainable Electrical Energy.

TenneT, 2009. Quality and Capacity Plan 2008–2014.