Essays in Oligopoly Theory

by

NICOLAE GABRIEL DRĂGAN

B.Sc. (Eng. Physics), The University of Bucharest, 1992

A Thesis Submitted in Partial Fulfilment of
the Requirements for the Degree of

Doctor of Philosophy

in

The Faculty of Graduate Studies

(Department of Economics)

We accept this thesis as conforming
to the required standard

The University of British Columbia

June 2003

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Economics

The University of British Columbia
Vancouver, Canada

Date June 19, 2003

DE-6 (2/88)

Abstract

This thesis comprises three essays that analyze some strategic interactions of firms in an oligopolistic setting.

The first essay examines the strategic role of partial commitments. Allowing firms to deviate from their announcements at some cost relaxes the usual assumption of irrevocable commitments. Such flexible commitments are shown to increase competition and welfare when the firms' actions are strategic substitutes (quantities), but to facilitate collusion and decrease welfare when they are strategic complements (prices). Assuming flexibility to be purely of a technological nature, the model is extended to allow firms to choose between two technologies characterized by different marginal costs of production and flexibilities. With strategic substitutes, adoption of an inflexible and cost-inefficient technology can lead to increased competition and welfare.

The second essay investigates the role of capacities in sustaining collusion in a repeated game. It is shown that collusion can be supported by choosing insufficient capacities if ex ante identical firms tacitly agree to produce the same output regardless of their ex post capacities. This type of self-enforcing agreement removes the rent-seeking consequences of the usual type of tacit agreements used in the literature. An exogenous increase in the discount factor is shown to possibly increase welfare by improving the cost-efficiency of firms. An exogenous increase in the number of firms may lead to an increase in the cost-inefficiency of firms and, therefore, it may decrease welfare.

The last essay examines the effect of entry in one regional market on the structure and competition in another regional market that interacts with the first one through a firm (a national firm) that operates in both markets. If entry changes the national firm's profitability of undertaking activities that reduce its marginal cost, possible entry in one regional market will impinge on the profitability of the firm in the other regional

market. It is shown that entry in one market can either intensify or reduce competition in the other regional market. One surprising result is that an attempt by government to subsidize entry in one market may motivate the national firm to deter entry in both markets, but this does not necessarily reduce welfare.

Contents

List of Figures

Acknowledgements

I gratefully acknowledge the help and suggestions offered by Mukesh Eswaran. I also wish to thank Margaret Slade and Guofu Tan for helpful discussions. I also benefited from discussions with Matthew Aharonian, Gorkem Celik, Ron Cheung, Patrick Francois, Parikshit Ghosh, Francisco Gonzalez, Stephanie McWhinnie, Rob Petrunia, Okan Yilankaya and other participants of the microeconomic theory workshop at UBC. None of this would have been possible without the support of Roxana Drăgan.

CHAPTER 1

Introduction

This thesis comprises three essays that analyze some strategic interactions of firms in an oligopolistic setting. I examine how market structure and performance are determined in three scenarios: those involving partial commitments, capacity constraints and demand-independent markets, respectively. Since their effects are important and sometimes counterintuitive, I argue that they are worth investigating.

Chapter 2, entitled 'Flexible Commitments', examines the strategic role of commitments that can be revoked at a non-prohibitive cost (I will refer to this type of commitments as *partial* or *flexible* commitments). The majority of the sequential move models used in the literature assume perfect commitments, from which no player can deviate at any cost. This is often an inconvenient assumption because the order of moves comes from modelling considerations and it could be hard to defend in empirical studies. This difficulty is exacerbated by the fact that order of moves can have a large impact on firms' profits, and therefore this extreme assumption leads to the question of how the interactions between them generates the first/second mover advantage. Historical accidents may explain the order of moves in some cases, but the assumption of such accidents is ad hoc.

One way to avoid this difficulty is to endogenize the leadership decision, by adding a pre-play stage in which firms decide who moves first (Hamilton & Slutsky 1990) or by allowing firms to change their minds as often as they want by paying a cost that increases as the time approaches a deadline (Caruana & Einav 2002).

The main purpose of using partial commitments is not to solve the leadership question, but to soften the impact of order of moves on firms' profits, and therefore making an assumption about 'historical accident' more palatable. In addition to this, an obvious

advantage of partial commitments is that they are more realistic than perfect commitments. A more subtle advantage is that strength of commitment can itself become a strategic choice variable. For example, if different production technologies offer a different commitment abilities, then firms will take this into account when deciding which technology to adopt.

The base model I construct considers two firms that sequentially announce their intended future actions, and then simultaneously choose their actual actions. Similar to the approach taken by Henkel (2002), a firm here can deviate from its announcement by paying a deviation cost. For a given deviation, the penalty to be paid captures the extent to which the announcement represents a commitment.

If firms play a quantity game, flexible commitments facilitate competition: both firms aggressively use the commitment ability to increase their outputs, leading to a lower equilibrium price. I show that, surprisingly, *both* firms (including the first mover) would rather prefer to be unable to commit at all than to have low strength partial commitments available. In the case in which firms play a price game, partial commitments have an opposite effect: they facilitate collusion. If the commitments are weak, the first mover has the opportunity to signal its willingness to collude by announcing a high price, while it still retains the ability to charge an actual low price if the rival does not follow suit. *Both* firms can benefit from the ability to commit.

The last part of the Chapter 2 analyzes the issue of technology choice when commitment strength is technological in nature. Flexible technologies are usually praised because they allow to firms to adapt to demand shocks, while the fact that they could offer different commitment strengths is largely ignored. The appearance and adoption of flexible manufacturing systems in the last part of the twentieth century marks this as a particularly important issue. More recently, the appearance of 'smart labels' promises to further enhance the flexibility of those systems of production. Other sources of flexibility can be procurement practices (buying resources in large batches vs. continually buying small batches) or labor contracts (bilateral vs. multilateral labor contracts).

To analyze this issue I assume there are two technologies: an old technology, which is both inflexible (offers high commitment ability) and inefficient (high marginal cost of production), and a new technology, which is flexible and has a low marginal cost of production. Firms simultaneously pick a technology, and then, given their technologies, they play the base model with quantities.

An improvement in the flexible technology (reduction in its marginal cost) could induce one firm to switch from the inflexible to the flexible technology and, surprisingly, the equilibrium price could increase. This is due to the fact that the flexibility precludes the firm from credibly committing to a higher output, and therefore, the competition softens to the extent that the total welfare decreases. This contradicts the usual presumption that adoption of more efficient technology decreases prices.

Being in a better position to exploit the ability to commit, the first mover always picks a technology that is at least as inflexible as the technology chosen by the second mover.

Chapter 3, entitled 'Supporting collusion with insufficient capacity', examines the role played by capacity in enforcing tacit collusion. The traditional view maintains that excess capacity is inversely correlated with cartel stability, because firms with excess capacity can profit more from deviating. The modern view, convincingly argued by Benoit & Krishna (1987), states that excess capacity helps tacit collusion in a supergame, because firms are better prepared to punish 'rogue' firms. It is quite surprising that the empirical support for the modern view is weak at best. In this chapter I argue for the restoration of the traditional view, using a non-standard tacit agreement in a setup similar with that used by Benoit & Krishna (1987).

Interpreting one input (capital) of a Cobb-Douglas production technology as capacity, I consider a tacit agreement (called *egalitarian*) in which both firms agree to produce the same output, regardless of how much capacity they install ex-post. I show that when firms non-cooperatively chose capacities, knowing that this choice is to be followed by a infinitely repeated competition in quantities, they choose to support tacit collusion

by installing less capacity than the level that would minimize the cost of producing the equilibrium output.

The usual result of excess capacity of the modern view derives from the fact that with the usual tacit agreement the firms with higher capacity are rewarded with a higher share of cartel's profit. Therefore, excess capacity is just an outcome of rent-seeking behaviour. Firms hold excess capacity and cannot obtain monopoly profit even if they manage to maintain the monopoly quantity and price. With the egalitarian agreement there is insufficient capacity and, given a high discount factor, the cartel can achieve the output, price and *profit* of a monopolist.

The main advantage of the egalitarian agreement is its simplicity (to detect cheating a firms just has to estimate competitors' output). In addition of this, given a high discount factor, this agreement allows firms to obtain higher profits than with the usual tacit agreement. Its main disadvantage, I demonstrate, is that it is not self enforcing for low discount factors.

Firms install sub-optimal capacities as a signal to show their willingness to collude. Therefore, anything that helps them collude decreases their need to restrain their capacity and makes them more cost-efficient. On one hand, an increase in the discount rate facilitates collusion because future penalties have a higher weight but, on the other hand, firms curtail their capacities by less and become more efficient, thus potentially leading to an increase in welfare. Similarly, an exogenous increase in the number of firms hinders tacit collusion and firms might find it necessary to increasingly curtail their capacities and this productive inefficiency can decrease total welfare. When the number of firms is large, however, a further increase in the number improves welfare. Therefore, an entry promoting policy cannot be guaranteed to be welfare enhancing unless it manages to induce a large number of firms to enter the market.

Chapter 4, entitled 'Competition in regional and national markets', is an essay investigating the interaction between firms in two separate markets (markets with unrelated demands). The link between these markets is the fact that only one firm (called *national* firm) competes in both markets with regional firms. The national firm is the only one

able to undertake R&D activities to decrease its marginal cost of production in both markets. Entry in one market affects the national firm's profitability of undertaking R&D activities , and therefore affects the profitability of the firms in the other market.

Note that firms use constant return to scale production functions, and therefore the link between markets is not caused by a production technology displaying increasing marginal cost of production, like in the analysis of Bulow, Geankopolos & Klemperer (1985). For markets linked by rising marginal costs, Chen & Ross (2002) compare the interactions between a national firm and regional firms to the interactions between two national firms in multimarket contact.

The scenario I investigate could apply to industries like air transport (national carrier competing with local carriers), telecommunications (a firm like Telus competing as local exchange carrier in both BC and Alberta), or internet service providers (national providers competing with many local providers in many geographical areas). It has a special relevance to the Canadian airline industry, which became very concentrated in the aftermath of deregulation.

The essential ingredient of the model is the national firm's ability to undertake *one* type of R&D activity that reduces its marginal cost of production in *both* regional markets.

The term 'regional' should not be interpreted only in its geographical meaning: the analysis presented here it is likely to be valid for two differentiated products that share a common stage of production. For example a regional market could be the market for operating systems for desktop computers, overwhelmingly dominated by Microsoft with its range of Windows 9x/ME/XP products, and the second regional market can be considered the market for operating systems for servers, where Microsoft competes with its Windows XP/2000/NT against Sun's Solaris and Linux, among others.

There is a fundamental tendency for entry to reduce the profitability of R&D activities and, therefore, to increase the marginal cost of production. But with sufficiently cheap R&D technology, entry in one market makes firms in the other market worse off when there is a small number of firms in the entry market. In this case the price in the

entry market falls more than in the other market: the price falls in both markets because the national firms is a tougher competitor (lower marginal cost) but it falls more in the entry market because of increased competition (larger number of firms). As a result, the national firm's behaviour might seem predatory to new entrants.

If the number of firms is endogenous and the national firm is allowed to deter entry, it can do so by decreasing its marginal cost to the point where its rivals' profit becomes zero. The decision to deter entrants in a market is tightly connected with what happens in the other market. If regional markets have different costs of entry, the national firm might tolerate entrants in the market with low cost of entry, because it can exploit the market with high cost of entry. If entry cost decreases in the high cost market, then this market becomes less profitable to the national firm. The national firm might be induced to deter entry in both market because the sacrifice required to obtain this is smaller. Surprisingly, a policy aiming to attract entry in the high cost market by subsidizing entry (tax exemptions, for example) can result in deterrence in *both* regional markets. Alternatively, if the national firm deters entry in both markets, a decrease in the cost of entry in the *low* cost market can induce the national firm to accommodate entry in both regional markets.

The welfare results are ambiguous: entry deterrence decreases industry's profit, but increases consumer surplus because of the lower prices required to keep the other firms out of the market.

CHAPTER 2

Flexible Commitments

I analyze the strategic effects of partial commitments (proposed actions that can be revoked by paying a penalty) in a duopoly in which firms sequentially announce their proposed actions, and then simultaneously choose their actual actions. I show that, in the case of strategic substitutes (outputs), partial commitments increase competition and welfare, and that it is possible for both firms to be worse off than in the no commitment case. In the case of strategic complements (prices), partial commitments decrease welfare and can facilitate collusion. Extending the model by adding an initial stage in which firms simultaneously choose from technologies characterized by different commitment strengths, I show that the first mover always chooses a technology at least as inflexible as the one picked by the second mover when higher inflexibility is associated with a higher marginal cost of production. Furthermore, I show that the introduction of more efficient technologies can be welfare-reducing.

1. Introduction

That commitments are irrevocable is usually the maintained assumption in sequential move models. Not only is this assumption often unreasonable (one can break a contract at a cost that is not always prohibitive), but it also has a large impact on the equilibria emerging in such models. The most direct way of relaxing the assumption of perfect commitment is to allow players to deviate from their proposed actions by incurring a penalty. In this paper I will argue that the degree of difficulty of deviating from an announced action can play a major strategic role in oligopoly and, therefore, that it requires a careful analysis. In particular, the welfare effects of the introduction of a technology need to be judged not only by its production efficiency, but also by the

commitment ability it offers. I will show that it is possible for a cost-inefficient technology to be welfare enhancing, if it offers stronger commitment potential that leads to increased competition.

Since the order of moves in a game strongly influences firms' payoffs, one difficulty with using irrevocable commitments is that the researcher has to carefully justify the order of moves in his model. Sometimes it is relatively easy to observe whether sequential or simultaneous moves are appropriate to a certain phenomenon, but it is often the case that the time structure is not obvious. For example, if competitors observe each other's moves with a large delay, then a sequential move model would seem appropriate, whereas if firms cannot observe competitors' moves, a simultaneous move model would be the appropriate choice. It is unclear what order of moves should be assumed when the observation delay is not large.

One way to resolve this difficulty is to allow players to endogenize the leadership decision. Players might have the same preference over the order of moves, but prefer not to move simultaneously. For example, Hamilton & Slutsky (1990) endogenize the time structure in a base game by adding a pre-play stage in which two players simultaneously choose whether to enter early or late in the base game. If both choose to enter at the same time, then the basic model is a simultaneous moves game; otherwise the player choosing to enter early becomes the leader in a sequential moves base game. For this type of games, where the base game is a quantity duopoly, Amir & Grilo (1999) provide sets of minimal conditions on the demand and cost functions for which a simultaneous or sequential equilibrium emerges.

Even if models like those alluded to above might explain, for example, why a certain player becomes the leader when both players can and prefer to move first, there is another problem that has to be addressed in a sequential move game. As Henkel (2002) observes, the sequential structure assumes irrevocable decisions. In particular, the leader can completely commit to a certain action that can never be changed, at any cost. In practice, commitments are not likely to be completely irrevocable. A player can deviate

from his previously announced action, but he has to pay a penalty. For example, a firm can contract a large amount of inputs as a commitment to produce a large output. Such

a contract may be renegotiated later, but would require reimbursing the suppliers for disrupting their plans.

Henkel (2002) studies the impact of partial commitments on equilibrium payoffs and allows the degree of commitment to be endogenously determined in a three stage model. In the first stage, player 1 simultaneously chooses the degree of commitment (i.e. the penalty he has to pay in case of deviation) and announces the action he intends to play in stage three. In the second stage, player 2 irreversibly chooses an action. In stage three, player 1 chooses his actions (which can be different from the action previously announced in stage one) and the payoffs are realized. Player 1 becomes a leader if he chooses a very large penalty cost, and he becomes a follower by choosing a zero deviation cost. Henkel proves that in the case of strategic substitutes player 1 chooses to be a leader, obtaining the Stackelberg leader's payoff. But in the case of strategic complements he might obtain, by partially committing, a payoff above that of a Stackelberg follower. The intuition behind the second result is that player 1 signals his intention to sustain a collusive action (say, high price), but he cannot fully commit since he has to be able to punish an aggressive action of the second player. Note that, by assumption, player 2 is in fact perfectly committed to his action, even though the payoffs are realized later.

In Henkel (2002), in addition to the assumed asymmetric time structure, firms are also asymmetric in their abilities to commit: player 1 can choose any degree of commitment, but player 2 is completely bound to his choice, even though payoffs are realized later. In my analysis I remove the asymmetry in players' commitment ability. My model also entails three stages: in stage one, Firm 1 announces the action it intends to play in stage 3; in stage two, Firm 2 announces the action it intends to play in stage 3; in stage three, firms simultaneously choose their actions and payoffs are realized. If actions differ from announcements, each deviating firm has to pay a penalty. I examine the impact of exogenous commitment strength on the subgame perfect equilibrium actions and payoffs.

I then extend this base model by adding a stage zero in which firms simultaneously choose between two technologies with different commitment strengths and marginal

costs. I use this extension to study the trade-off between the strategic value of commitment strength and the advantage of having more efficient production technologies.

If penalty costs are infinite, stage three of the model becomes irrelevant from a strategic point of view, since firms are completely bound by their announcements, and the game reduces to one of Stackelberg competition. If penalty costs are zero, stages one and two have no strategic value since announcements cannot influence the stage three competition, and the model reduces to a symmetric, simultaneous moves model. In this game, increasing the penalty cost from zero to infinity changes Firm 1 from a symmetric simultaneous move duopolist into a Stackelberg leader. In contrast, in Henkel's (2002) model player 1 changes from a follower to a leader. In particular, in my setup, Firm 1 always realizes either a higher or a lower profit than Firm 2, and the profit ranking does not change with commitment strength.

The results depend on whether player's actions are strategic substitutes or strategic complements. In a linear demand and constant marginal cost setup, I use quantity competition to model the strategic substitutes case and price competition to model the strategic complements case.

The results for the strategic substitutes case differ from those of Henkel (2002): in my model, while Firm 1 would prefer irrevocable commitments, its profit is not monotonic in commitment strength (as measured by the penalty for deviating). Firm 1 would rather prefer no commitment at all to commitments of moderate strength. Moderate commitment strength gives incentives to both firms to increase their output, but does not allow Firm 1 to sufficiently exploit its announcement to compel Firm 2 to decrease its output in equilibrium. On one hand, Firm 2 would like to decrease its output after Firm 1 announces a high output, but, on the other hand, the ability to partially commit gives incentives to Firm 2 to increase its output relative to the no commitment case. Firm 1 foresees this and realizes that its own announcement will be not as successful in inducing the other firm to decrease its output (relative to the case where Firm 2 cannot commit), and therefore decides to announce a lower output. The overall result is that *both* firms increase their output, though Firm 1 increases it by more than Firm 2.

The results for the case of strategic complements resemble those of Henkel (2002). In particular, for moderate commitment strength, Firm 1 obtains a profit higher than a Stackelberg leader's. This arises because Firm 1 can announce, say, a friendly price while leaving itself flexible enough to fight a price war if Firm 2 announces a price that is too low.

Another difference between the strategic substitutes and the strategic complements cases is that, while commitments increase competition with substitutes, they serve as a collusion-facilitating device with complements. Commitments increase welfare in the case of strategic substitutes, but they decrease consumer surplus and also total welfare in the case of strategic complements.

In the extended model, in which firms choose between two technologies, the firm moving first always chooses a technology that is at least as inflexible as the one chosen by the second mover. This offers one plausible explanation for the fact that quite often the incumbent firms are perceived to be less flexible than the new entrants in the market.

The welfare impact of choosing technologies is surprising in the case of strategic substitutes. When technologies that offer more production efficiency are available, firms may choose these over inefficient technologies that offer greater scope for commitment. Since commitment opportunities can increase competition, the availability of more efficient technologies can generate scenarios in which competition is diluted to the extent that welfare is reduced, despite the lower production costs. This possibility — which I demonstrate can obtain as an equilibrium outcome — calls into question our common assumption that cost reductions lead to price reductions. What is missing in the logic of this inference is the notion that the commitment possibilities that technologies offer may have a bearing on the degree of competition that prevails. Inefficient technologies that allow greater scope for commitments can increase competition and enhance welfare.

These results are sensitive to the manner in which the penalty cost increases with deviations from announcement. In particular, linear and constant penalty cost allow for stronger commitments, and all the non-monotonicity in commitment strength exhibited in the case of quadratic costs disappears. In the linear case, increasing commitment

strength changes the model monotonically from a symmetric, simultaneous move game to a Stackelberg game. With linear or constant deviation cost, deviations are never observed in equilibrium, whereas they invariably are in the quadratic deviation cost case.

The finding that results depend on assumed functional forms suggests that general results are not to be had in models of this nature. The paper, however, is intended to communicate that there are certain outcomes supported by robust economic intuition which undermine some very basic notions that we routinely accept. One such notion is the common belief that more efficient technologies will lower prices.

2. The Model

There are two firms in the market, labeled Firm 1 and Firm 2. Both firms have zero marginal cost of production. The model has three stages:

(i) in stage 1, Firm 1 makes an announcement, S_1, regarding the action it will choose in stage three.

(ii) in stage 2, Firm 2 makes an announcement, S_2, regarding the action it will choose in stage three.

(iii) in the third stage, firms compete by simultaneously choosing actions s_1 and respectively s_2. Sales take place and firm i realizes a gross profit $\pi_i(s_i, s_j; S_i, S_j)$, with $i = 1, 2$ and $i \neq j$. If $s_i \neq S_i$, then Firm i has to pay a penalty $c(S_i - s_i)^2$, with $c > 0$.

The parameter c is interpreted as commitment strength, and it is exogenous to the model.

Probably the easiest way to rationalize the cost of deviating from the announcement is to assume that firms use contracts with suppliers of inputs as announcements and these contracts can be observed by rivals. For example, a firm contracting a large amount of inputs announces it intention to produce a high level of output. If the firm decides to produce a different quantity, it will have to renegotiate the contract and reimburse suppliers for the disruption it causes. Since the cost function is quasi-convex

in quantities, suppliers will agree to a renegotiated contract if the reimbursement is quasi-convex in deviation from the initial contract. This would suggest that a quasi-convex deviation cost[1] is reasonable. This section will present the case of strictly convex deviation cost where, for analytical tractability, I assume quadratic costs. Section 3 presents the case of linear deviation costs.

Firm i chooses an announced action and actual action pair, (S_i,s_i)[2]. Since the announcement S_i can affect the best response schedule of Firm i, it can also affect the Nash equilibrium of the stage three game. It is reasonable to suspect that, in spite of having to pay a penalty for deviation, a firm might find it profitable to deviate from its announcement, if this deviation leaves it in a better position for the last stage competition.

The pure strategy subgame perfect equilibrium of this game is solved for in the usual manner, by using backward induction. First the Nash equilibrium of the stage 3 game is solved to find the equilibrium actions $s_i^*(S_i, S_j)$, given the announcements. The equilibrium profit of Firm 2 can be written as:

$$\pi_2^*(S_1, S_2) = \pi_2 \left[s_1^*(S_1, S_2), s_2^*(S_1, S_2); S_1, S_2\right] - c\left[s_2^*(S_1, S_2) - S_2\right]^2$$

The optimal announcement of Firm 2, denoted by $S_2^*(S_1)$, maximizes its profit and therefore solves the equation[3]:

$$\frac{\partial \pi_2^*(S_1, S_2)}{\partial S_2} = 0$$

Now the profit of Firm 1 can be rewritten as:

$$\pi_1^*(S_1) = \pi_1 \left[s_1^*(S_1, S_2^*(S_1)), s_2^*(S_1, S_2^*(S_1)); S_1, S_2^*(S_1)\right] - c\left[s_1^*(S_1, S_2^*(S_1)) - S_2^*(S_1)\right]^2$$

[1]Actually, this suggests that the penalty cost is quasi-convex in the deviation from the contracted quantity of input. With a linearly homogenous production technology, the penalty cost is also quasi-convex in the deviation from the 'announced' output quantity.

[2]The strategies of Firm 1 would be $\{S_1,\ s_1(S_1, S_2)\}$ and the strategies of Firm 2 would be $\{S_2(S_1),\ s_2(S_1, S_2)\}$. The announcement of Firm 2 is generally contingent on Firm 1's announcement and the final action of each firm is generally contingent on the announcements of both firms.

[3]With the associated second order condition $\frac{\partial^2 \pi_2^*(S_1,S_2)}{\partial S_2^2} \leq 0$. This condition, as well as the second order conditions for the other maximization problems in this paper, is satisfied.

Finally, the optimal announcement of Firm 1, denoted by S_1^* solves the equation:

$$\frac{\partial \pi_1^*(S_1)}{\partial S_1} = 0$$

Since the case of strategic substitutes assumes a different demand function than the case of strategic complements, and each case yields different qualitative results, I separate their analysis. The following subsection considers the strategic substitutes case, while the next one deals with strategic complements.

2.1. Strategic Substitutes. In this case, assume firms play a game in quantities and quantity announcements. I will use q_i, Q_i instead of the generic notation s_i, S_i, with $i \in \{1, 2\}$.

The inverse demand function is $p = 1 - Q$, where Q is the total amount sold in the market, $Q = q_1 + q_2$.

The profit of firm i is:

$$\pi_i(q_i, q_j, Q_i) = (1 - q_i - q_j)q_i - c(q_i - Q_i)^2 \quad , \tag{2.1}$$

where $i, j \in \{1, 2\}$ and Q_i is the announcement of firm i. First, we have to solve the Cournot-Nash equilibrium of the last stage of the model.

The best response function of Firm i, $r_i(q_j, Q_i)$, is:

$$r_i(q_j; Q_i) = \frac{1 - q_j + 2\,c\,Q_i}{2 + 2\,c} \tag{2.2}$$

The equilibrium quantities of the third stage of the game,denoted by q_i^*, solve the system of equations $q_i^* = r_i(q_j^*; Q_i)$, with $i, j \in \{1, 2\}$ and can be written as:

$$q_i^*(Q_i, Q_j) = \frac{1 + 2\,c + 4\,c\,(1 + c)\,Q_i - 2\,c\,Q_j}{3 + 8\,c + 4\,c^2} \tag{2.3}$$

The output of Firm i increases in its announcement and decreases in that of the other firm. If Firm i announces a large output, it becomes less profitable for it to produce less than that, because of the deviation penalty. If Firm j ($j \neq i$) announces a large output, shifting out its own reaction function, then increasing q_i becomes less profitable, and Firm i will produce less. This gives an advantage to Firm 1, which is the first to announce, and by announcing a high output, it constrains the scope of Firm 2's

announcement. Nevertheless, Firm 2 is still capable of increasing its output, but less so than Firm 1.

The first two stages of the game can be solved in the manner presented in the beginning of this section, and the subgame perfect equilibrium quantities, denoted by q_i^C, are:

$$
\begin{aligned}
q_1^C(c) &= \frac{27 + 150\,c + 348\,c^2 + 416\,c^3 + 256\,c^4 + 64\,c^5}{81 + 432\,c + 928\,c^2 + 1008\,c^3 + 560\,c^4 + 128\,c^5} \\
q_2^C(c) &= \frac{\left(3 + 8\,c + 4\,c^2\right)\left(9 + 26\,c + 24\,c^2 + 8\,c^3\right)}{81 + 432\,c + 928\,c^2 + 1008\,c^3 + 560\,c^4 + 128\,c^5}
\end{aligned} \tag{2.4}
$$

Figure 1 illustrates the equilibrium output levels in the SPNE as functions of c. The output of the Firm 1 increases with c, as expected, but for low commitment strength, the output of Firm 2 also increases with c. Low commitment strength prevents the leader from credibly increasing its actual output by enough (via large announced outputs) to overcome follower's incentives to increase output.

This result is summarized by the following proposition, which is proved in Appendix 6:

PROPOSITION 2.1. *$q_1^C(c)$ is monotonically increasing with c and there is a critical value of c, denoted by c_1, such that $q_2^C(c)$ increases with c for $c < c_1$ and $q_2^C(c)$ decreases with c for $c > c_1$.*

The subgame perfect equilibrium profits of each firm are presented in Figure 2. The profit of the follower monotonically decreases with c but, for low commitment strength, the leader's profit may also decrease with c.

Denoting by $\Pi_i^C(c)$ the equilibrium profit of firm i, this result is summarized by the following proposition, which is proved in Appendix 6:

PROPOSITION 2.2. *$\Pi_2^C(c)$ is monotonically decreasing with c and there is a critical value of c, denoted by c_2 such that $\Pi_1^C(c)$ decreases with c for $c < c_2$ and $\Pi_1^C(c)$ increases with c for $c > c_2$.*

A direct corollary of Proposition 2.2 is that for low commitment strength, both duopolists are worse off than in the case in which no commitment is possible($c = 0$).

It is well known that such an outcome can obtain in games in which commitments are chosen simultaneously (e.g. Brander & Spencer 1983). The availability of a vehicle of commitment makes it individually rational, in these models, to use these commitments despite the fact that it is collectively detrimental. It is not appreciated, however, that this outcome could obtain in games with sequential moves, too, where Firm 1 is endowed with a first-mover advantage.

The next step of the analysis concerns the total welfare generated by the industry. The total equilibrium output and therefore consumer surplus increases with c, but industry profit decreases with c for low values of c, raising the question of whether total surplus increases with c. In the absence of income effects, one may consider the area under the demand curve and above the price line as the consumer surplus and, therefore, the total surplus (as the sum of producer surplus and consumer surplus) becomes:

$$TS = \Pi_1^C + \Pi_2^C + CS = \Pi_1^C + \Pi_2^C + \frac{1}{2}Q^2 \tag{2.5}$$

where $CS = \frac{1}{2}Q^2$ is the consumer surplus, with Q being the total output sold in the market.

The following proposition, proved in Appendix 7.1, summarizes the welfare effects of an increase in c:

PROPOSITION 2.3. *With strategic substitutes, the total surplus is monotonically increasing in c.*

Commitment possibility offers incentives to *both* firms to increase their output, pushing down prices and increasing consumer surplus. Appendix 7.1 offers the formal proof of this claim.

If firms could exercise influence over the strength of commitment and strong commitments are infeasible, in view of Proposition 2.2, they both would seek to decrease c and this would decrease the welfare generated by the industry.

2.2. Strategic Complements. For this case I assume firms play a game in prices and prices announcements, and therefore the generic notation used at the beginning of this section changes from s_i, S_i to p_i , P_i, with $i \in \{1,2\}$.

The demand functions are:

$$q_i = 1 - p_i + bp_j \quad , \tag{2.6}$$

where p_i is the price charged by firm i, $0 < b < 1$ and $i, j \in \{1,2\}$.

The profit of Firm i becomes:

$$\pi_i(p_1, p_2, P_i) = p_i(1 - p_i + bp_j) - c(p_i - P_i)^2 \quad , \tag{2.7}$$

where P_i is the announcement of Firm i.

When the value of c increases from 0 to infinity, the equilibrium in this model gradually departs from a symmetric, Bertrand equilibrium towards a Stackelberg equilibrium in prices.

First we have to solve the stage three game in prices.

The best response of firm i, $r_i(p_j; P_i)$, is:

$$r_i(p_j; P_i) = \frac{1 + b\,p_j + 2\,c\,P_i}{2 + 2\,c} \tag{2.8}$$

The Bertrand-Nash equilibrium prices, denoted by p_i^*, solve the system of equations:$p_i^* = r_i(p_j^*; P_i)$, with $i, j \in \{1,2\}$ and can be written as:

$$p_i^*(P_i, P_j) = \frac{2 + b + 2\,c + 4\,c\,(1+c)\,P_i + 2\,b\,c\,P_j}{4\,(1+c)^2 - b^2} \tag{2.9}$$

The crucial difference between the above equation and equation (2.3) for quantities is that here a high price announcement of firm i gives incentives to *both* firms to charge higher prices. This immediately raises the question of whether a high price announcement can serve as an instrument that facilitates collusion.

Stages two and one of the model are easily solved following the procedure outlined at the beginning of this section.

Denoting by $p^B(c)$ the equilibrium prices, numerical simulations show that:

PROPOSITION 2.4. *$p_1^B(c)$ is monotonically increasing with c and there is a critical value of c, denoted by c_3 such that $p_2^B(c)$ increases with c for $c < c_3$ and $p_2^B(c)$ decreases with c for $c > c_3$.*

For small values of c, the commitment is a collusion-facilitating instrument: Firm 1 announces a friendly (i.e. high) price, but it can easily deviate to punish unfriendly announcement of Firm 2, and this enables both of them to raise prices. Once the commitment becomes strong, Firm 1 loses the ability to punish, and Firm 2 will be able to offer even lower prices. This explains the hump-shaped profile of Firm 1's equilibrium price as a function of c.

For the case $b = \frac{1}{2}$, Figure 3 presents the dependency of the subgame perfect equilibrium prices on commitment strength, while Figure 4 presents the subgame perfect equilibrium profits for each firm.

Denoting by Π_i^B the equilibrium profit of Firm i, one can numerically check the following proposition:

PROPOSITION 2.5. *$\Pi_2^B(c)$ monotonically increases with c and there is a critical value of c, denoted by c_4, such that $\Pi_1^B(c)$ increases with c for $c < c_4$ and $\Pi_1^B(c)$ decreases with c for $c > c_4$.*

Figure 4 depicts the above proposition for the case $b = \frac{1}{2}$.

Note that Firm 1 would like an intermediate value of c, because it can obtain a higher profit than that of the price-Stackelberg leader (which in turn is higher than the Bertrand equilibrium profit.) To reach this outcome, Firm 1 needs both flexibility and commitment: the commitment enables it to credibly signal its willingness to maintain a high price, while flexibility enables it to punish an aggressively low announcement by Firm 2. If commitment strength increases sufficiently, Firm 2 can announce a low price without fear of harsh competition in stage three.

Partial commitments negatively affect the welfare generated by the industry. Industry profit increases with c, but the consumer surplus (measured in by the expenditure

function) decreases. The consumer surplus part of welfare dominates, and total surplus decreases in c. Appendix 7.2 presents a way to set up the welfare function.

3. Linear deviation cost.

Commitment strength cannot be captured perfectly by a single parameter like c, because it depends also on the functional form characterizing the deviation penalty.

The results of section 2 partly change if the deviation penalty is assumed to be linear, and not quadratic, in deviation. That is, the penalty cost is $c|S_i - s_i|$. For the case of strategic substitutes, the equilibrium quantity of Firm 2 then decreases monotonically in c. In the quadratic cost case, as we have seen, Firm 2's equilibrium output decreases only for strong commitment strength. Firm 1 is hurt at low c because weak commitments are not successful in inducing Firm 2 to decrease its output by much. But with linear deviation cost, Firm 1's commitment is strong enough to convince the competitor that it is unprofitable to increase output. This suggests the conjecture that the linear deviation cost specification strengthens the commitment.

This conjecture turns out to be indeed correct and the reason is that the marginal cost of deviation is constant for linear cost, whereas it is near zero for small deviations when deviation cost is quadratic. Assume for the moment that Firm 2 produces q_2 in the third stage game and Firm 1 is on its best reply function for the case $c = 0$, and Firm 1 announced exactly this level of output, $(Q_1 = q_1)$ in stage one. Since the announcement $Q_1 = q_1$ reduces the deviation penalty to zero, q_1 is also on Firm 1's best reply function when $c > 0$ for the announcement $Q_1 = q_1$. Mathematically, this is described by $q_1 = r_1(q_2; Q_1) = r_1(q_2; q_1)$. What is the best reply of Firm 1 if Firm 2 produces a slightly lower output, $q_2 - dq_2$, given the announcement $Q_1 = q_1$? Without commitment, the best reply of Firm 1 would be to slightly increase its output to $q_1 + dq_1$. Since at $q_1 + dq_1$ Firm 1's revenue is maximized, the marginal revenue is zero and so deviating to q_1 decreases revenue by an insignificant amount. With linear deviation cost, the marginal cost of deviating from announcement is c (strictly greater than zero), and therefore, by sticking with the announcement $Q_1 = q_1$, Firm 1 obtains significant

deviation cost savings with an insignificant revenue loss. Since the marginal revenue is always lower than c everywhere between q_1 and $q_1 + dq_1$ (for small dq_1), the highest profit is realized at q_1. Mathematically, this can be written as $q_1 = r_1(q_2 + dq_2; Q_1)$, for any small $dq_2 \geq 0$. Graphically, the best reply of Firm 1 has a vertical portion at the announced level of output, as it shown in Figure 5A. The best reply of Firm 1 for an announcement Q_1 consists of a segment along the line R^H (the reaction function for a large announcement) up to Q_1, a vertical segment between lines R^H and R^L (the reaction function for a zero output announcement) at Q_1 and a segment along line R^L starting Q_1.

This contrasts with the quadratic cost case. As above, the marginal revenue is zero at the point (q_1+dq_1, q_2+dq_2), but the marginal deviation cost is also zero at the announced output level. Firm 1 does not find it profitable to stick with the announcement because it saves a negligible amount in deviation cost, but it loses a small but strictly positive revenue by doing so. The optimum actual output level is somewhere between q_1 and q_1+dq_1, at the point where the marginal deviation cost equals the marginal revenue. The range of values of q_2 for which Firm 1 finds it optimal to stick with the announcement is a singleton when the deviation cost is quadratic, but it has an infinite number of points when the deviation cost is linear. Alternatively, it can be said that changing the announcement affects the reaction function only locally when the deviation cost is linear, but globally when the deviation cost is quadratic. The vertical distance between lines R^H and R^L increases with the commitment strength, c. The shape of the best response function when deviation cost is linear is that used by Dixit (1980) in his well-known capacity commitment model. Because of this similarity, the rest of this exercise is somewhat reminiscent of the analysis of Ware (1984), who extended Dixit's study in a similar fashion by allowing for commitments also by the second mover.

As it can be seen from Figure 5, by changing Q_1 Firm 1 can cover all points between lines R^H and R^L. Firm 2 also has analogous reaction functions. Graphing the set of points covered by both reaction functions in Figure 5B, we see that the intersection between best reply functions of Firm 1 and 2 can occur only in the cross-hashed, diamond

shaped area. This area is the area in which the Nash equilibrium of the stage three game can take place, and it is magnified in Figure 7.

We can now characterize the solution to the third stage of the game in the linear deviation cost case. The following lemma is useful for the next part of the exercise.

LEMMA 3.1. *With linear deviation costs, no deviations from announcement are observed in the subgame perfect equilibrium of the game.*

PROOF. Assume that the subgame perfect equilibrium strategies are $\{(Q_1^e, q_1^e), (Q_2^e, q_2^e)\}$, with $Q_2^e \neq q_2^e$. Since it is a Nash equilibrium for the stage three game, it must be the case that Firm 2's reply function and Firm 1's reply function intersect at (q_1^e, q_2^e). Because Firm 2's reply function has a flat segment at its announced output, the Nash equilibrium can also be reached with an announcement $Q_2 = q_2^e$. The revenue would be the same as in the assumed SPNE, but Firm 2 would realize a higher profit because it would not have to pay any deviation penalty. Therefore, $\{(Q_1^e, q_1^e), (Q_2^e, q_2^e)\}$ cannot be a SPNE if $Q_2^e \neq q_2^e$. Exactly the same reasoning applies for Firm 1, provided that a change in Firm 1's announcement does not induce Firm 2 to change its announcement and, therefore, its reaction function. If $Q_1^e \neq q_1^e$ then one of the two cases (A and B) depicted in Figure 6 must occur. (Note that Firm 2 always picks the most extreme available isoprofit contour curve that is feasible, because it offers the highest profit). In case A, the announcement is lower that the actual output. Increasing Q_1 towards q_1^e does not induce Firm 2 to change its announcement and therefore leaves the revenue unchanged, but Firm 1's profit increases because of lower deviation penalty. Therefore, Figure 6A cannot depict a SPNE. In case B, the announcement is higher than the actual output and decreasing Q_1 towards q_1^e does not induce Firm 2 to change its announcement, until the vertical segment of Firm 1's reply function hits the westmost isoprofit contour of Firm 2. Once the vertical segment hits this contour, the isoprofit contour moves with the vertical segment and it remains tangent to it. To reach it, Firm 2 has to pick a reaction function with a flat segment that intersects the vertical part of Firm 1's reply function at its tangency point with the

above mentioned westmost isoprofit contour. In other words, in Figure 6B, decreasing the announcement of Firm 1 either does not induce any changes in Firm 2's behavior, or induces Firm 2 to change its announcement such that its flat part intersects the vertical part of Firm 1's reply function. So the situation illustrated in case B of Figure 6 cannot be an SPNE, either. □

Graphically, Lemma 3.1 says that in the subgame perfect equilibrium the vertical segment of the Firm 1's best response function and the flat segment of Firm 2's reaction function must intersect.

A graphical way of finding the subgame perfect equilibrium is presented in Figure 7, which details a region around the symmetric, simultaneous Nash equilibrium ($c = 0$), depicted by point E'. The line going through points A and B (B and C) graphs the best response function of Firm 1 (Firm 2) when it announces a large output. The line going through points D and C (A and D) graphs the best response function of Firm 1 (Firm 2) when it announces a small output. Point B (point D) is the Nash equilibrium of the third stage if both firms announce large (small) outputs. Point A is the Nash equilibrium if Firm 1 announces a large output and Firm 2 announces a small output, and point C depicts the Nash equilibrium when Firm 1 makes a small announcement while Firm 2 makes a large announcement. The two lines going through point E' are the best response functions when there is no cost to deviation ($c = 0$). The 'vertical' dashed curves are isoprofit contours of Firm 2, with the profit on a contour being lower than the profit on a contour to the west of it. The solid 'horizontal' curve line is an isoprofit contour of Firm 1. The coordinates of a point X in this region ($X \in \{A, B, C, D, E, E'\}$) are labeled (q_1^X, q_2^X).

According to Lemma 3.1, the subgame perfect equilibrium (if it exists) has to be in region ABCD, since that is the only region in which the respective flat, and vertical, segments of the best response functions can meet. Firm 1 prefers the southern part of this region, while Firm 2 prefers the western part of it.

Consider the highest feasible isoprofit contour of Firm 2, which is therefore tangent to the line AB at point T. Consider point E, which is the point at which this isoprofit

contour is tangent to the vertical portion of Firm 1's reply function, and also the intersection of the thick solid and dashed reply functions in the picture. I claim that this point E characterizes the subgame perfect equilibrium of the game.

Suppose first that Firm 1 announces an output higher than q_1^E. Firm 2 tries to reach the highest feasible isoprofit contour. That is, Firm 2 maximizes its profit by choosing a point on Firm 1's reaction line that is inside region ABCD. In this case point T maximizes its profit and Firm 2 can reach it by announcing q_2^T. The equilibrium would take place at the point T, where Firm 2's reaction function is tangential to the downward part of Firm 1's reaction function. This contradicts Lemma 3.1 and, therefore, Firm 1 cannot announce an output higher than q_1^E in the subgame perfect equilibrium.

What happens if Firm 1 announces an output lower than q_1^E? Firm 2 reaches the maximum profit where line EE' meets the vertical part of Firm 1's reaction function. The equilibrium would be at a point on line EE', to the northwest of point E.

If c is small, then the isoprofit contour of Firm 1 is relatively flat at point E and line EE' is steeper than this isoprofit contour at point E. Therefore, all points on the EE' line to the northwest of point E are above Firm 1's isoprofit curve going through point E, and Firm 1 obtains a lower profit than it can realize at point E. Announcing an output lower than q_1^E does not maximize Firm 1's profit and it cannot be a part of an SPNE strategy.

If c increases, the distance between opposite sides of the diamond ABCD increases: line AB shifts out and the highest feasible isoprofit contour of Firm 2 moves to the east. This enables Firm 1 to increase its announcement, and the equilibrium moves southeast from point E, along the line EE'. This happens until the Stackelberg equilibrium is reached: at that point Firm 1's isoprofit line that goes through point E becomes tangent to line EE', and moving further south-east on line EE' would actually decrease Firm 1's profit.

Denoting by $q_i^L(c)$ the subgame perfect equilibrium output of Firm i, this result is summarized by the following proposition:

PROPOSITION 3.2. *With linear deviation cost and strategic substitutes, there is a critical value of c, denoted by c_L, such that in the subgame perfect equilibrium:*

(i) *$q_1^L(c)$ monotonically increases and $q_2^L(c)$ monotonically decreases with c for $c < c_L$, and*

(ii) *$q_1^L(c)$ and $q_2^L(c)$ are constant and reach the Stackelberg output of the leader and follower, respectively, for $c \geq c_L$.*

Since point E in Figure 7 is always on Firm 2's reply function without commitment, Firm 2 remains on the same reaction function, independently of commitment strength.

This does not mean that Firm 2 does not take advantage of the commitment possibility. Firm 2's access to partial commitment helps it to partially compensate for the fact that it is a follower: without commitment, the reaction function of Firm 2 would be the line EE' in Figure 7 and the equilibrium would be at the intersection of lines EE' and AB, where Firm 2 would produce less that in the case in can partially commit. This is analogous to the outcome in Ware's (1984) analysis of the model of Dixit (1980), when the second mover is also allowed to commit to capacity.

The main difference between the linear and quadratic cost is that the non-monotonic behavior with respect to c disappears when deviation cost is linear in deviation. In particular, the first mover always prefers a commitment of any strength to the no commitment situation. Another difference is that the Stackelberg equilibrium is reached for a finite value of c in the case of linear cost, but it requires perfect commitment ($c = \infty$) in the case of quadratic cost.

This section easily applies for constant deviation costs, where a deviation of any size imposes a penalty that is independent of the extent of the deviation. Since the marginal cost of deviation from announcement is infinite, the reaction functions have a flat/vertical segment. Similar intuition applies, and similar results obtain: firms do not deviate from their announcements and the model moves from a simultaneous move game to a Stackelberg game when the deviation cost increases from zero to infinity.

4. Choosing flexibility

In this section, I consider the issue of how firms would choose between technologies that offer different degrees of flexibility (that is, have different values of c). I assume the source of flexibility is explicitly technological in nature. For example, an old iron melting plant is likely to offer lower flexibility with respect to output than a modern iron melting micro-plant. A firm with good logistics can reduce inventories, decrease the marginal cost of production and be more flexible than a firm that does not invest much to improve its logistics.

There are many reasons why a firm would like to become flexible (for example to better react to demand shocks), but, in my setup here, flexibility comes mainly as a by-product of the technology. Of course, improving the flexibility of a technological process can be interpreted as generating a new technology. I am considering only the case of strategic substitutes. Here flexibility is not directly desired: a flexible firm is at a strategic disadvantage because it cannot credibly commit to a large level of output. To balance this effect, I assume flexibility is indirectly desired, because it is associated with lower marginal cost of production. So better commitment ability comes at the price of greater production cost.

From now on I assume that firms have available only two technologies. The term 'old technology' denotes the technology with low flexibility and high marginal cost, while 'new technology' denotes the one with high flexibility and low marginal cost.

I consider the strategic substitutes case with quadratic deviation cost presented in Section 2.1 but in which a technology has a constant marginal cost of production m_T, where T designates the production technology, old (O) or new (N). The cost penalty of deviating from announcement Q_i is $c_T(Q_i - q_i)^2$, where q_i is the actual sold output, and c_T is interpreted as a measure of inflexibility of the technology T. The assumptions regarding the old and new technology can be written as:

$$\begin{aligned} c_O &\geq c_N \\ m_O &\geq m_N \end{aligned} \tag{4.1}$$

With an inverse demand function $p = 1 - q_1 - q_2$, the profit of Firm i using technology T, π_{iT}, can be written as:

$$\pi_{iT} = (1 - q_1 - q_2 - m_T)q_i - c_T(Q_i - q_i)^2 \tag{4.2}$$

with $i \in \{1, 2\}$ and $T \in \{O, N\}$.

The timing structure of the model is as follows:

(0) In stage 0, both firms simultaneously choose their technologies.

(1) In stage 1, Firm 1 makes an announcement, Q_1, about the quantity it intends to produce in stage 3.

(2) In stage 2, Firm 2 makes an announcement, Q_2, about the quantity it intends to produce in stage 3.

(3) In stage 3, firms simultaneously choose their actual outputs, and profits are realized.

The subgame perfect equilibrium is solved for by backward induction. In stage 3 the Nash Cournot equilibrium is solved for, given technologies and announcements. In stage 2, Firm 2 finds its optimal announcements, given technologies and Firm 1's announcement, while in stage 1 Firm 1 computes its optimal announcement given technologies. This is done precisely in the manner presented in Section 2. In stage 0, the firms simultaneously choose their technologies with full awareness of the ensuing outcomes in announcements and outputs.

Label by $\pi_i(T_1, T_2)$ the stage 0 profit of Firm i in the ensuing subgame perfect Nash equilibrium when Firm 1 uses technology T_1 and Firm 2 uses technology T_2. Since in this stage there are only four possible outcomes, I will numerically investigate each of them separately as possible Nash equilibria. For simplicity, I will focus on the case where $c_N = m_N = 0$, and I will discuss later the case where $c_N > 0$. This implies that, while the new technology produces output costlessly, it affords no opportunity for commitment.

Once the Nash equilibrium choices in stages three, two and one are solved for backwards, given the firms' technology choices, the technology choice problem reduces to

a simultaneous move game with two strategies (O and N) for each firm, with payoffs $\pi_i(T_1, T_2)$. The normal form of this game is presented in Table 1.

		Firm 2: O	Firm 2: N
Firm 1	O	π_1(O,O), π_2(O,O)	π_1(O,N), π_2(O,N)
	N	π_1(N,O), π_2(N,O)	π_1(N,N), π_2(N,N)

TABLE 1. Normal form of the technology choice game

The purpose of the remaining exercise is to find under which conditions (values of m_O and c_O) a cell of Table 1 represents a Nash equilibrium.

Some explanation for a quick understanding of the Figures used in the following subsections:

- The marginal cost of the old technology, m_O, is measured on the horizontal axis. Since m_N is assumed zero, m_O can be interpreted as the production cost disadvantage of the old technology.
- The degree of inflexibility of the old technology, c_O, is measured on the vertical axis. As above, it can be interpreted as the inflexibility of the old technology relative to the new one.
- The thick curves labelled IC_1 are the loci of points in (m_O, c_O) space for which, in the SPNE of the continuation game, Firm 1 is indifferent between the old and the new technology for a given technology of Firm 2. Along this curve, Firm 1 is indifferent between two vertically adjacent cells in Table 1.
- Similarly, the thin curves labelled IC_2 are the loci of points in (m_O, c_O) space for which, in the SPNE of the continuation game, Firm 2 is indifferent between the old and the new technology for a given technology of Firm 1. Along this curve, Firm 2 is indifferent between two horizontally adjacent cells in Table 1.

4.1. Firm 1 with old technology, Firm 2 with new technology. If Firm 1 is to use the old technology when Firm 2 uses the new, it must be the case that

$$\pi_1(O, N) \geq \pi_1(N, N). \tag{4.3}$$

In Figure 8, the region where this is satisfied is represented by the area to the left of the thick line IC_1, along which Firm 1 is indifferent between the two technologies. To maintain its indifference, a higher marginal cost of the old technology needs to be compensated for by a greater commitment advantage (that is, greater inflexibility). This is why the indifference locus of Firm 1 is positively sloped always (even for the other three cases considered below). Above this locus, the commitment advantage of the old technology (because of high c) is so large that Firm 1 strictly prefers that technology. The marginal profitability of inflexibility is decreasing in c_O and this explains the convex shape of the solid line on which the incumbent is indifferent between the old and the new technologies.

Similarly, for Firm 2 to choose the new technology when Firm 1 picks the old, it must be the case that

$$\pi_2(O, N) \geq \pi_2(O, O). \tag{4.4}$$

The region where this is satisfied is represented by the area to the right of the thin line in Figure 8. Along this IC_2 'indifference' curve, Firm 1 uses the old technology. For low inflexibility, increasing m_O makes the old technology more attractive to Firm 2: the low inflexibility does not allow Firm 1 to increase its output by much, and therefore there is scope for Firm 2 to exploit the commitment possibility. Therefore, to keep the new technology as profitable as the old one when m_O increases, its cost advantage has to increase too, and the thin curve is upward sloping. At a sufficiently high level of inflexibility, Firm 1 becomes so privileged in the first mover position that Firm 2, as the second mover, becomes unable to exploit its commitment ability adequately. (After all, with both firms using the old technology and high inflexibility, the model comes arbitrarily close to replicating Stackelberg competition, where it does not matter whether the second firm can commit; any additional increase in c_O just increases the first

mover's advantage and the second mover's disadvantage.) This makes the old technology less attractive to Firm 2 and therefore, to keep Firm 2 indifferent, the production cost of the old technology has to decrease when c_O increases, and the thin curve becomes downward sloping.

The area where both of the above conditions are satisfied is the hashed area between the thick and thin lines. In this area Firm 1 prefers to stick with an inefficient technology because it offers a greater strategic advantage, and Firm 2 prefers the new technology because of its lower production cost.

What happens with Firm 2 along the line $m_O = m^*$ in Figure 8, when m_O is small and Firm 1 uses the old technology ? If c_O is small, Firm 1's first mover advantage can induce Firm 2 to restrict output only slightly. Because of its cost advantage, Firm 2 would prefer to use the new technology. As c_O increases, the greater inflexibility of the old technology allows Firm 1 to credibly commit to higher levels of output, reducing the scope of Firm 2's production cost advantage. What happens if Firm 2 switches to the old technology? The thin line shows that it might be more profitable for Firm 2 to give up the small cost advantage and to exploit the stronger commitment offered by the old technology. When c_O increases further (above IC_2), the strategic advantage of Firm 1 increases (ultimately it becomes a Stackelberg leader) and the benefit Firm 2 extracts from being able to commit decreases until the cost advantage of the new technology becomes the overwhelming reason to switch back to the new technology.

Recall that I had normalized c_N to zero. Increasing c_N to a moderate level rotates the thin curve to the left and the solid curve to the right: the region in parameters space where the (O,N) equilibrium obtains expands. Increasing c_N to a high level rotates/shifts both curves to the left, contracting the region in which the (O,N) equilibrium obtains.

4.2. Firm 1 with new technology, Firm 2 with old technology. If Firm 1 is to use the new technology when Firm 2 uses the old, it must be the case that

$$\pi_1(N, O) \geq \pi_1(O, O). \tag{4.5}$$

In Figure 9, the region where this is satisfied is represented by area to the right of the IC_1 curve.

Similarly, for Firm 2 to choose the old technology when Firm 1 picks the new, it must be the case that

$$\pi_2(N,O) \geq \pi_2(N,N). \tag{4.6}$$

The region where this is satisfied is represented by area to the left of the IC_2 curve in Figure 9. Since these two regions do not overlap, this case cannot obtain. If Firm 2 finds it profitable to use the old technology, then Firm 1, which is better positioned to exploit inflexibility, would use the old technology, too.

4.3. Both firms with new technology. If Firm 1 is to use the new technology when Firm 2 uses the new, it must be the case that

$$\pi_1(N,N) \geq \pi_1(O,N) \tag{4.7}$$

In Figure 10, the region where this is satisfied is represented by area to the right of the IC_1 curve .

Similarly, for Firm 2 to choose the new technology when Firm 1 picks the new, it must be the case that

$$\pi_2(N,N) \geq \pi_2(N,O) \tag{4.8}$$

The region where this is satisfied is represented by area to the right of the IC_2 curve in Figure 10. Since we have assumed $c_N = 0$, the new technology offers no commitment opportunity. Thus, regardless of which firm announces output first, in this case the firm using the old technology is de facto the first mover. Thus $\pi_1(N,N) = \pi_2(N,N)$ and $\pi_1(O,N) = \pi_2(N,O)$. It follows that the two 'indifference curves' IC_1 and IC_2 coincide.

If the production cost advantage of the new technology is large, then both firms will use it regardless of how inflexible the old technology is. After all, the maximum profit a firm can get with the old technology is the profit of a Stackelberg leader, and that can be brought arbitrarily close to zero by increasing m_O.

4.4. Both firms with old technology. If Firm 1 is to use the old technology when Firm 2 uses the old, it must be the case that

$$\pi_1(O,O) \geq \pi_1(N,O) \tag{4.9}$$

In Figure 11, the region where this is satisfied is represented by area to the left of the IC_1 curve.

Similarly, for Firm 2 to choose the old technology when Firm 1 picks the old, it must be the case that

$$\pi_2(O,O) \geq \pi_2(O,N) \tag{4.10}$$

The region where this is satisfied is represented by area to the left of the IC_2 curve in Figure 11 (since this condition is the reverse of that for case (O,N), the thin line is identical with the one in Figure 8). In the shaded region, both firms use the old technology in equilibrium.

This last case somewhat resembles a prisoners' dilemma: as long as m_O is small and c_O moderate, both firms try to strategically exploit the old technology's inflexibility to the detriment of the other. When c_O is large, Firm 2 realizes it cannot profitably use commitments, and switches to the new technology. Note that this region shrinks, and may even completely disappear, if the new technology is not completely flexible (i.e. it disappears if we drop the $c_N = 0$ normalization and allow c_N to be large).

Figure 12, shows in one diagram the regions in parameter space that yield as equilibrium outcomes the technology choices (O,N), (N,O), (N,N) and (O,O). Observe that there are no 'holes' in the parameter space, that is, every pair (m_O, c_O) belongs to at least one of the technology regimes: (O,O), (O,N), or (N,N). In other words, there always exists a (pure strategy) subgame perfect equilibrium in technology choices. Furthermore, there is no overlap in the regions; with the exception of the borders between

them, they are mutually exclusive. In other words, except along the borders, there is a unique pure strategy SPNE associated with each (m_O, c_O) pair[4].

It can be shown that for no SPNE to exist, it must be the case that Firm 2 must benefit by switching to the old technology while Firm 1 does not benefit from such a switch[5]. Since Firm 1, the first mover, is in a better position to exploit inflexibility with strategic substitutes, I find that this condition is not satisfied.

4.5. Flexibility choice: Short summary and welfare analysis. Figure 12 can be used to summarize the results of the exercise. To the left of the thin line, the new technology does not emerge in equilibrium. Between the thin and thick lines, Firm 1 uses old and Firm 2 uses new technology. To the right of the full line, the new technology is so cost advantageous that it will be the only one emerging in equilibrium.

Examination of the shaded regions of Figure 12 reveals that the first mover firm will always choose a technology that is at least as inflexible as that of the second mover. Flexibility, therefore, is more likely a characteristic of followers.

The existence of non-zero measure area (O,O) in Figure 12 reveals that it is possible that both firms will choose the cost inefficient technology even when there is a more efficient technology available.

Defining total surplus as in equation (2.5) it is interesting to see what are the welfare changes along the border between region OO and ON (ON and NN) brought about by Firm 2 (Firm 1) switching from the new to the old technology.

Numerical computations show that everywhere along the OO–ON border, Firm 1's profit decreases discontinuously when Firm 2 switches from the new to the old technology. Firm 2's profit remains unchanged by this switch (this is the condition defining the border) and consumer surplus increases discontinuously. Consumer surplus always

[4]Along the boundaries between regions, one firm is indifferent between the two available technologies. Multiplicity of equilibria in this case can be avoided by simply assuming that whenever the two technologies are equally attractive to a firm, it chooses the old technology.

[5]A close examination of conditions (4.3)–(4.10) reveals that exactly four of them have to be satisfied. In particular, it must be the case that $\pi_1(O,O) \leq \pi_1(N,O)$ and $\pi_2(O,O) \geq \pi_2(O,N)$, or that $\pi_1(O,N) \leq \pi_1(N,N)$ and $\pi_2(N,O) \geq \pi_2(N,N)$.

increases by more than the decrease in Firm 1's profit, and the overall effect is *welfare enhancing.* Use of the inefficient technology increases welfare because the inflexibility associated with that technology induces an aggressive use of commitments and this increases competition.

Along the ON–NN border, by construction, Firm 1's profit does not depend on which technology it uses; however, Firm 2's profit decreases discontinuously when Firm 1 switches from the new to the old technology. Firm 1's ability to commit diminishes Firm 2's advantage of having a more efficient technology. At the same time, Firm 1 is able to increase its output, leading to an increase in consumer surplus. Once again, total surplus is increasing.

Any time a firm switches to the old technology (as a result of a small exogenous decrease in the cost advantage of the new technology), competition intensifies and the welfare generated by the industry increases.

Since equilibrium quantities and profits are continuous in c and m (as long as firms do not change their technologies), the above welfare effects work also for the case in which the old technology is in a small neighborhood to the east of the OO–NO or NO–NN borders. A decrease in the cost advantage of the new technology (caused, for example, by taxing the new technology) can increase welfare. Equivalently, an increase in the production cost of the new technology can increase welfare generated in equilibrium

The very real possibility alluded to above is worth emphasizing. That an exogenous increase in the cost efficiency of the new technology can bring about a switch to that technology by one of the firms is not surprising. What is surprising is that, in the new SPNE, the reduction in the opportunity to commit reduces the extent of the competition. Despite the decrease in production cost, the reduced competition brings about a reduction in welfare. That this emerges as an equilibrium outcome should question the very premise of our presumption that the introduction of more efficient technologies promotes competition. The degree of commitment possible with technologies should be an integral component of our thinking.

Eaton & Eswaran (1997) also have shown that firms might prefer an inefficient technology even when a more efficient one is available, if it helps them to collude better in an infinitely repeated game. In my setup, use of the old technology does not help firms to collude (there is no repeated game); it actually promotes competition.

5. Conclusions

Partial commitments allow for a finer assessment of the first mover advantage (or disadvantage). While the first mover always (i.e. regardless of commitment strength) enjoys a higher profit than the follower in a strategic substitutes game, it would rather prefer a no commitment situation if strong commitments are not feasible. The ability to commit gives incentives for both firms to increase their outputs, and therefore increases competition, potentially decreasing both firms' profits. Note that if the firm moving second is denied the ability to commit (as in Henkel (2002)), the first mover would always prefer a higher commitment strength. The results are different for a strategic complements case: Firm 1 would prefer a small commitment strength to both the no-commitment and strong commitment cases. My results for the strategic complements case is similar to that of Henkel (2002). Firm 1 values both commitment (because it helps the firm to credibly announce its intention to collude) and flexibility (since this enables it to threaten the other firm with vigorous competition in case it does not reciprocate with a similar announcement).

The welfare implications are quite different in the two cases: commitments stimulate competition in the strategic substitutes case, increasing welfare; and commitments promote collusion in the strategic complements case, decreasing welfare.

The ability to commit can make an inflexible and high production cost technology more attractive to both firms than a flexible one with low production cost. Even though the firm moving first is more likely to find the inflexible technology more profitable than the flexible one, it can also happen that both firms would prefer the cost-inefficient technology. The first mover always chooses a technology that is at least as inflexible as the technology picked by the second mover. The welfare results of firms' choice of

technology are surprising: whenever the two technologies have costs and flexibility such that a firm is indifferent between picking one or another, when it picks the cost-inefficient technology welfare increases because the higher inflexibility of that technology intensifies competition. An exogenous decrease in the attractiveness of the flexible technology (caused, for example, by a tax) can increase the welfare generated by the industry. Adoption of a technology that is cost inefficient but offers better commitment ability intensifies competition and it can compensate for the production inefficiency. The set of parameters for which the (exogenously induced) inflexible technology increase welfare has a measure larger than zero.

Appendix

6. Proofs for Propositions 2.1 and 2.2

Taking the derivative with respect to c in equations (2.4) immediately shows that $\frac{\partial q_1^C(c)}{\partial c} > 0$ for any positive c.

By taking the derivative of $q_2^C(c)$ one can easily work out that

$$sign(\frac{\partial q_2^C(c)}{\partial c}) = sign(243+540c-3288c^2-17280c^3-35264c^4-39232c^5-25088c^6-8704c^7-1280c^8)$$

Since the right hand side of the above equation has only one positive root, c_1. Because $\frac{\partial q_2^C(0)}{\partial c} > 0$ and there is only one positive root, it must be that $\frac{\partial q_2^C(c)}{\partial c} > 0$ for $c < c_1$, and $\frac{\partial q_2^C(c)}{\partial c} < 0$ for $c > c_1$. This concludes the proof for Proposition 2.1.

The SPNE profits for the two firms are:

$$\begin{aligned}
\Pi_1^C(c) &= \frac{(1+c)\left(3+6\,c+4\,c^2\right)^2}{81+432\,c+928\,c^2+1008\,c^3+560\,c^4+128\,c^5} \\
\Pi_2^C(c) &= \frac{(1+c)\left(9+26\,c+24\,c^2+8\,c^3\right)^2\left(9+32\,c+40\,c^2+16\,c^3\right)}{\left(81+432\,c+928\,c^2+1008\,c^3+560\,c^4+128\,c^5\right)^2}
\end{aligned}$$

Taking the derivative of $\Pi_2^C(c)$ immediately shows that Firm 2's profit decreases with c, for any positive c.

The derivative $\frac{\partial \Pi_1^C(c)}{\partial c}$ can be worked out to:

$$sign(\frac{\partial \Pi_1^C(c)}{\partial c}) = sign(-243-1152\,c-1260\,c^2+3168\,c^3+11520\,c^4+15872\,c^5+11712\,c^6+4608\,c^7+768\,c^8)$$

The right hand side of the above equation has only one positive root, c_2. Since $(\frac{\Pi_1^C(0)}{\partial c} < 0$ and there is only one positive root, it must be that $(\frac{\Pi_1^C(c)}{\partial c} < 0$ for $c < c_2$, and $(\frac{\Pi_1^C(c)}{\partial c} > 0$ for $c > c_2$.

7. Quadratic deviation cost: welfare

7.1. Strategic substitutes. Subgame perfect announcements, Q_1^C and Q_2^C of the two firms are:

$$Q_1^C = \frac{4\,(1+c)^2\,\left(9+30\,c+36\,c^2+16\,c^3\right)}{81+432\,c+928\,c^2+1008\,c^3+560\,c^4+128\,c^5} \tag{7.1}$$

$$Q_2^C = \frac{4\,(1+c)^2\,\left(9+26\,c+24\,c^2+8\,c^3\right)}{81+432\,c+928\,c^2+1008\,c^3+560\,c^4+128\,c^5} \tag{7.2}$$

and the industry output, Q, is:

$$q_1^C+q_2^C=\frac{54+300\,c+664\,c^2+736\,c^3+416\,c^4+96\,c^5}{81+432\,c+928\,c^2+1008\,c^3+560\,c^4+128\,c^5}. \tag{7.3}$$

The equilibrium profits, Π_1^C and Π_2^C, are:

$$\Pi_1^C = \frac{(1+c)\,\left(3+6\,c+4\,c^2\right)^2}{81+432\,c+928\,c^2+1008\,c^3+560\,c^4+128\,c^5}$$

$$\Pi_2^C = \frac{\left(9+26\,c+24\,c^2+8\,c^3\right)^2\,\left(9+41\,c+72\,c^2+56\,c^3+16\,c^4\right)}{\left(81+432\,c+928\,c^2+1008\,c^3+560\,c^4+128\,c^5\right)^2}$$

The total surplus defined by equation (2.5) works out to:

$$TS^C(c) = \frac{2\left(1458+15633c+75708c^2+218092c^3+414160c^4+542432c^5\right)}{\left(81+432c+928c^2+1008c^3+560c^4+128c^5\right)^2}+$$
$$+\frac{2\left(496992c^6+315072c^7+132480c^8+33408c^9+3840c^{10}\right)}{\left(81+432c+928c^2+1008c^3+560c^4+128c^5\right)^2}$$

It is easy to verify that $\frac{\partial TS(c)}{\partial c} > 0$. Therefore, total welfare is monotonically increasing with c, for the case of strategic substitutes.

7.2. Strategic complements. Because there are two interdependent markets, it is not possible to use the same procedure as in the case of strategic complements. With interdependent demands, integrating the areas to the left of the demand curves is dependent on the order of integration and therefore this procedure cannot yield a measure of consumer surplus. Since the proper consumer surplus is given by the expenditure function, the total surplus function can be defined as

$$TS^B(c,u;b)=\Pi_1^B(c;b)+\Pi_2^B(c;b)-E(c,u;b) \quad , \tag{7.4}$$

where Π_i^B is the equilibrium profit of Firm i ($i\in\{1,2\}$), and $E(c,u;b)$ is the expenditure required to reach utility level u at the equilibrium price vector $(p_1^B(c;b),p_2^B(c;b))$. An

increase in the expenditure function due to a change in c indicates that it is more expensive to reach utility u and thus consumer welfare decreases.

The utility function yielding the demand functions (2.6) is

$$U(q_1,q_2,q_3)=\frac{1+b}{1-b^2}\,(q_1+q_2)-\frac{b}{1-b^2}\,q_1\,q_2-\frac{1}{2\,(1-b^2)}\,({q_1}^2+{q_2}^2)+q_3 \tag{7.5}$$

provided that the consumption of the outside good, q_3, is strictly positive.

For the remaining part of this analysis I normalize the price of the outside good, p_3, to 1 and I assume that consumers' income is large enough to insure a strictly positive consumption of good 3.

The indirect utility function, $V(p_1,p_2,m)$ associated with the above utility function is

$$\begin{aligned} V(p_1,p_2,m) &= m-(1+b\,p_1-p_2)p2-(1-p_1+bp_2)p_1- \\ &-\frac{b}{1-b^2}\,(1+bp_1-p_2)(1-p_1+bp_2)+\frac{1+b}{1-b^2}(2-p_1-p_2+b(p_1+p_2))- \\ &-\frac{(1+bp_1-p_2)^2+(1-p_1+bp_2)^2}{2(1-b^2)}\quad , \end{aligned} \tag{7.6}$$

where m is the income of the consumer.

The corresponding expenditure function, $e(p_1,p_2,u)$ is

$$\begin{aligned} e(p_1,p_2,u) &= (1+bp_1-p_2)p_2+p_1(1-p_1+bp_2)+ \\ &+\frac{b}{1-b^2}(1+bp_1-p_2)(1-p_1+bp_2)-\frac{1+b}{1-b^2}(2-p_1+bp_1-p_2+bp_2)- \\ &-\frac{(1+bp_1-p_2)^2+(1-p_1+bp_2)^2}{2(1-b^2)}+u\quad . \end{aligned} \tag{7.7}$$

The expenditure function used in (7.4) is

$$E(c,u;b)=e(p_1^B(c;b),p_2^B(c;b),u).$$

Even though the reference utility u is exogenously given in formula 7.4, it does not play any significant role because the expenditure function (7.7) is additively separable in u. Therefore, the reference level of utility, u, does not matter when the total surplus

function is used for welfare ranking. Numerical simulations indicate that the total surplus monotonically decreases in c, for all values of b. Industry profit increases with c, but consumer surplus decreases even more. Even in the case of a large b, when the consumption of q_1 and q_2 is not very sensitive to an increase in *both* prices, consumer surplus still decreases by more than the increase in industry profit because the consumption of the outside good q_3 decreases sharply.

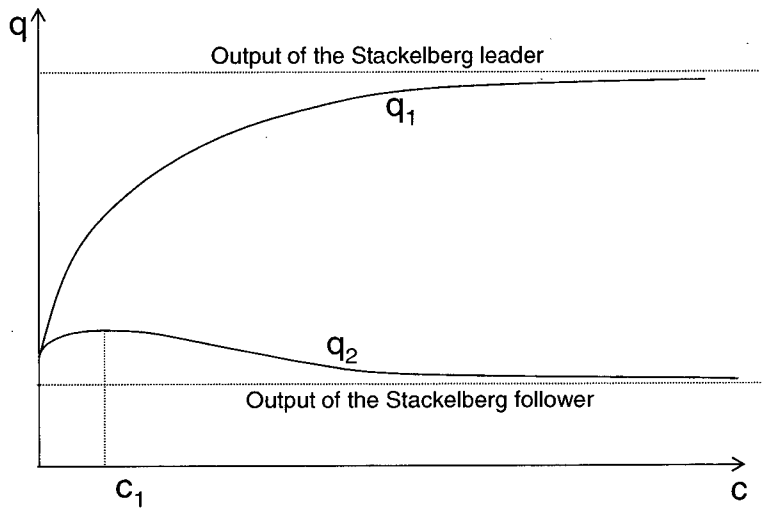

FIGURE 1. Strategic substitutes: subgame perfect equilibrium quantities.

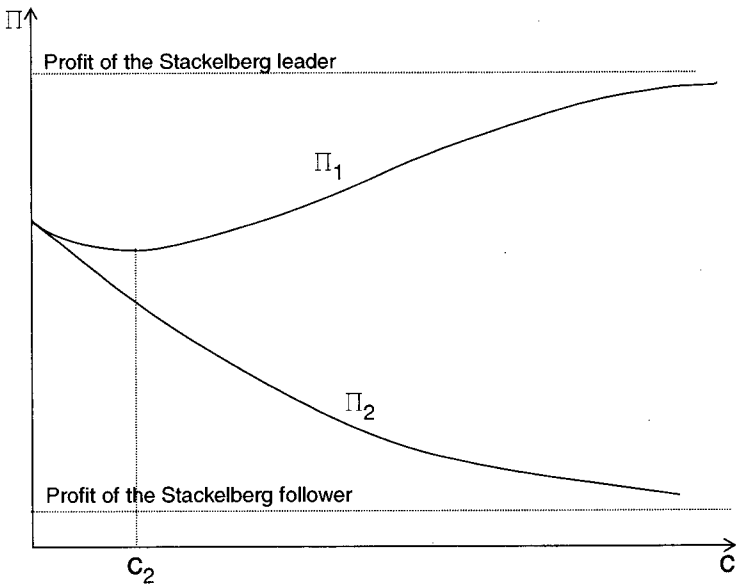

FIGURE 2. Strategic substitutes: subgame perfect equilibrium profits.

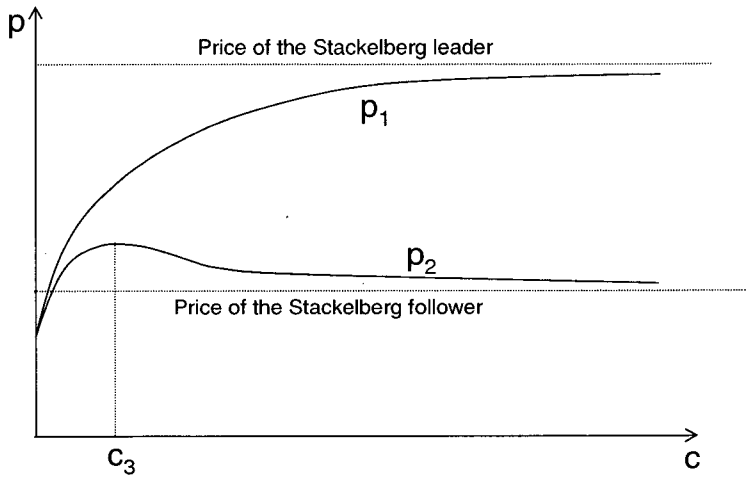

FIGURE 3. Strategic complements: subgame perfect equilibrium prices.

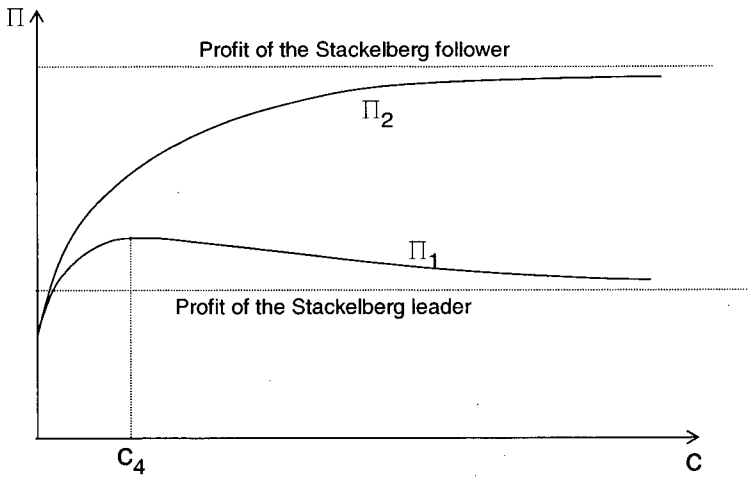

FIGURE 4. Strategic complements: subgame perfect equilibrium profits.

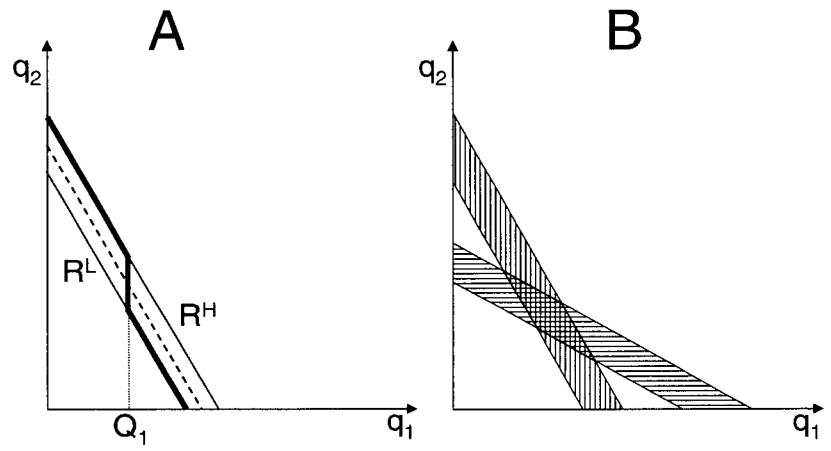

FIGURE 5. The stage three best response functions of Firm 1 and 2 in the case of linear deviation cost

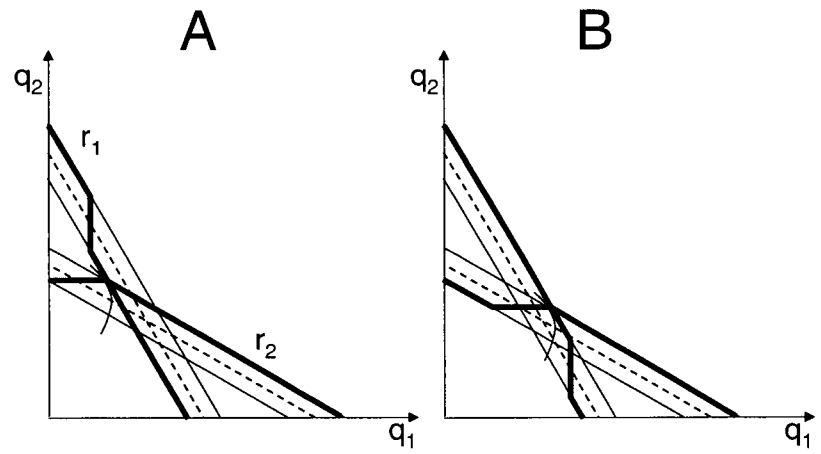

FIGURE 6. Firm 1 does not deviate from its announcement in the sub-game perfect equilibrium.

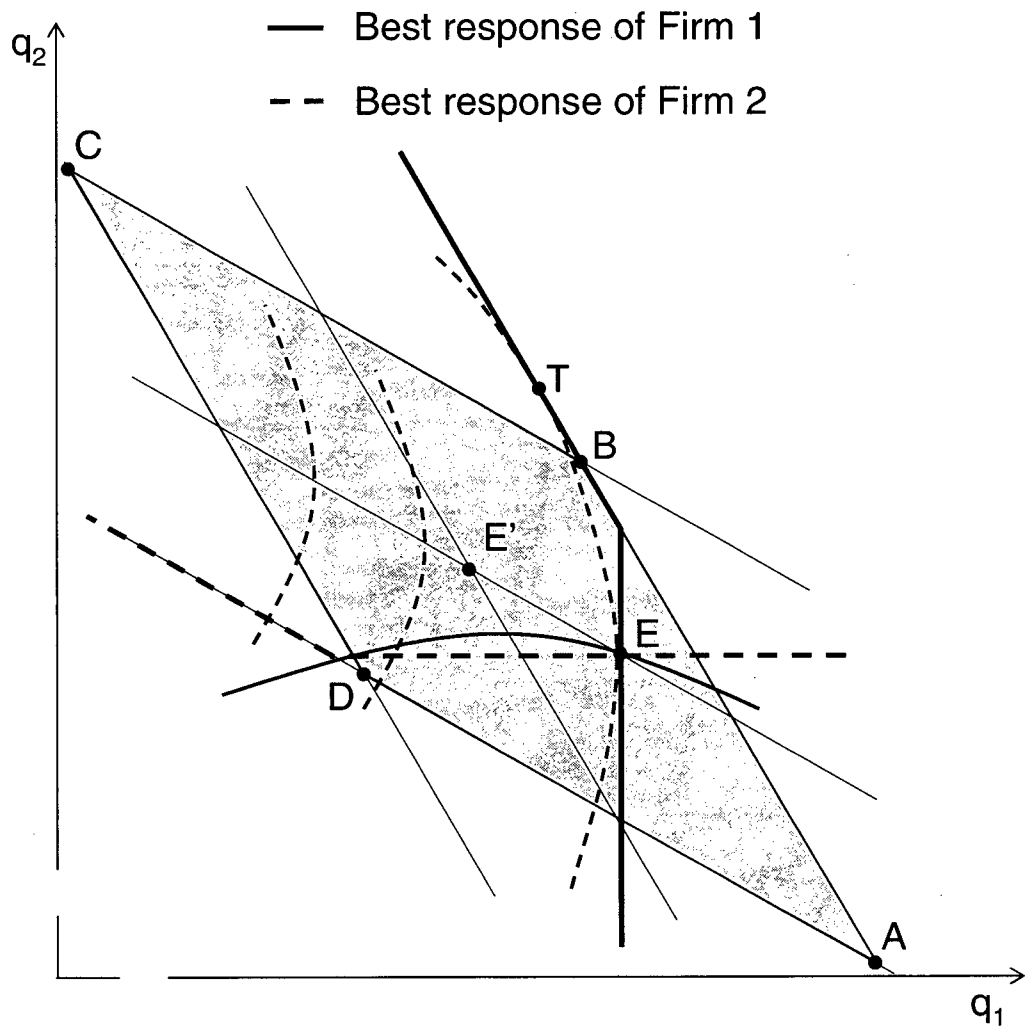

FIGURE 7. Subgame perfect equilibrium with linear deviation cost and small c.

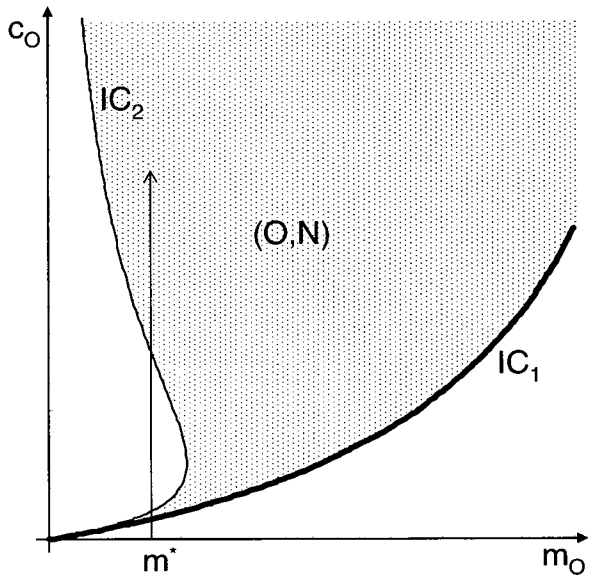

FIGURE 8. Subgame perfect equilibrium with Firm 1 using old technology and Firm 2 using new technology

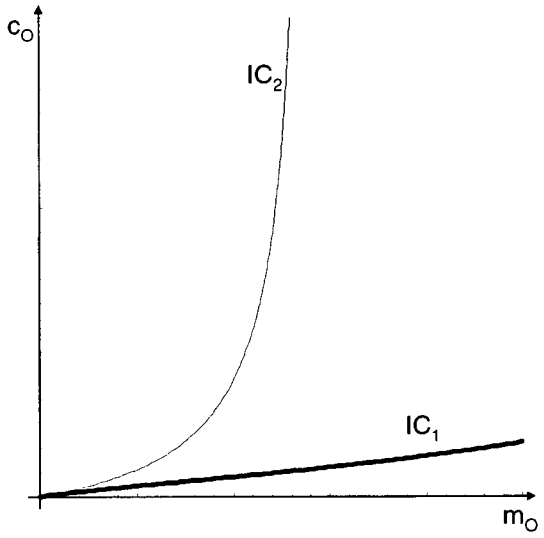

FIGURE 9. There is no subgame perfect equilibrium in which Firm 1 picks the new technology and Firm 2 chooses the old technology

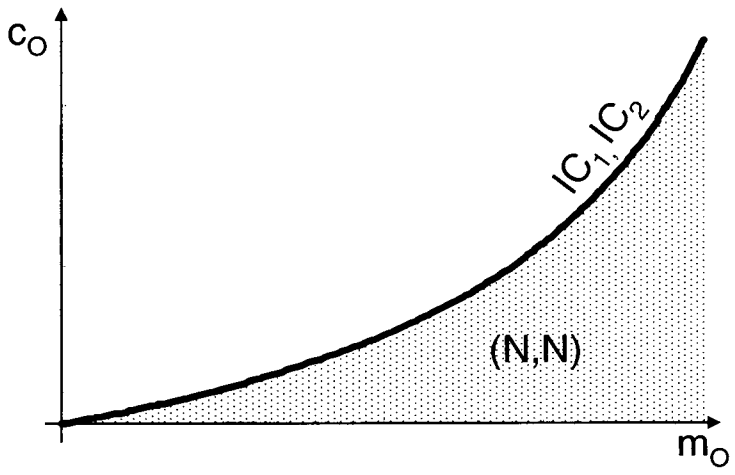

FIGURE 10. Subgame perfect equilibrium with both firms using the new technology

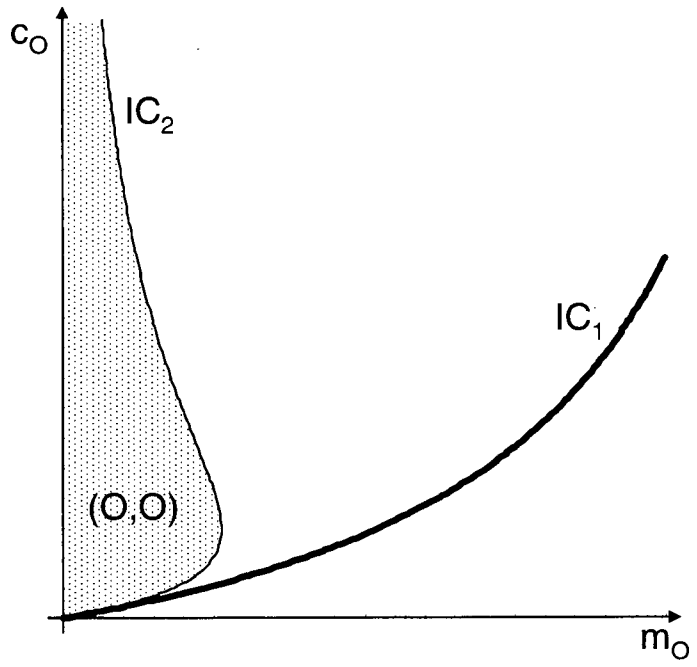

FIGURE 11. Subgame perfect equilibrium with both firms using the old technology

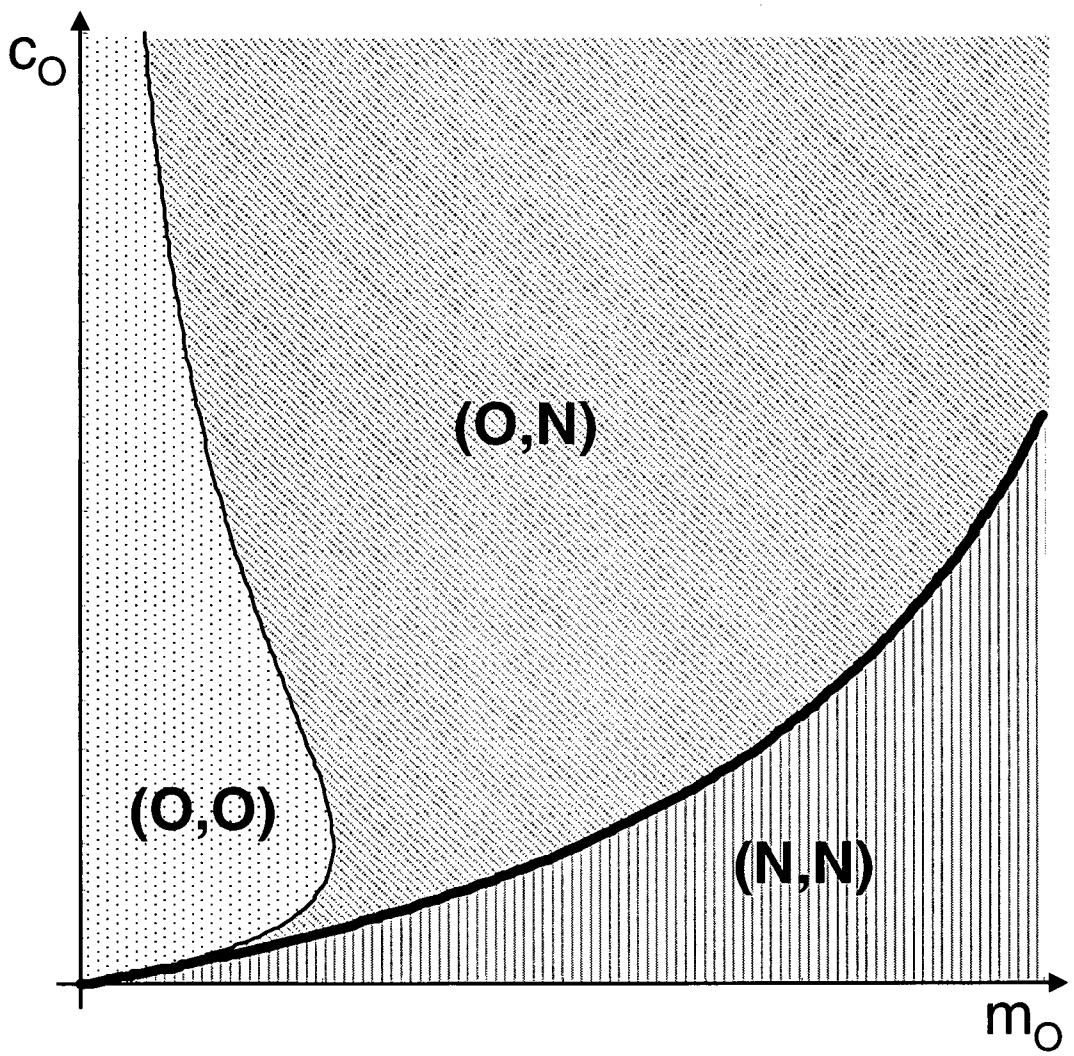

FIGURE 12. Partition of the (m_O, c_O) space according to the technologies adopted in the SPNE.

CHAPTER 3

Supporting collusion with insufficient capacity

This chapter investigates the role of capacity in supporting tacit collusion. I propose a self-enforcing tacit agreement in which ex-ante identical firms agree to produce the same output regardless of their capacities. This removes the rent-seeking consequences of the usual type of tacit agreements used in the literature. With a tacit agreement of this sort, I show that, in equilibrium, firms choose to facilitate collusion by installing insufficient capacity. As long as firms collude only partially, an exogenous increase in the number of firms leads to less efficient production (firms increasingly curtail their capacities), and may even decrease welfare. Likewise, an increase in the discount factor, despite facilitating collusion, may increase the welfare generated per period.

1. Introduction

Whether excess capacity helps or hinders tacit collusion in an oligopoly is an open question. The traditional view in the Industrial Organization literature is that cartel stability is inversely correlated with excess capacity, because individual cartel members can capture a greater market share by undercutting the price of rivals (Phlips 1995*a*, Scherer & Ross 1990). This view has been challenged by more recent game theoretical analyses, like those of Benoit & Krishna (1987) and Davidson & Deneckere (1990). They show that, in the context of repeated interactions, excess capacity facilitates collusion by providing individual members of a cartel greater ability to punish cheaters. In other words, the traditional view emphasizes that the instantaneous gain for a cheater increases with excess capacity, while the modern view recognizes that the long term punishment to cheating increases, too. In spite of these advances, the question is far from being settled and excess capacity never could be used as a evidence of collusion.

Empirically, too, the relationship between excess capacity and collusion is weak. As noted by Davidson & Deneckere (1990), the findings are rather inconclusive and most of the evidence supports the traditional wisdom. A representative example would be the cement and nitrogen fertilizer industry in Europe during the 1920s and 1930s. The firms colluded (sometimes even explicitly) during the 1920s; a capacity race took place during the collusion phase, only to lead to cartel collapse during the early 1930s[1]. Conlin & Kadiyali (1999) find in the Texas lodging industry that, while increased concentration of idle capacity may lead to lower prices, the effect of idle capacity on price is ambiguous. This ambiguity is also illustrated in practice by the U.S. Department of Justice and the Federal Trade Commission's 'Horizontal Merger Guidelines'. Here is it recognized that excess capacity in the hands of 'rogue' firms can increase competition, but in hands of 'non-rogue' firms can reduce competition and, therefore, excess capacity cannot be used as an indicator of collusion.

There are also recent studies that try to reconcile theory with the traditional view. Nocke (1999) uses firms' cartel participation constraints to derive a negative relationship between excess capacity and cartel stability[2], and Staiger & Wolak (1992) find that cartels break down[3] during periods of low demand (high excess capacity).

It is essential to clarify the concept of capacity used is this paper. In the usual[4] notion of capacity used in the literature, the marginal cost of production above capacity is essentially infinite. The most serious disadvantage of this concept is that it rules out the possibility of insufficient capacity in supporting collusion: in a simplified model with zero marginal cost of producing below capacity and costly capacity installation, by definition, firms will never produce less than the capacity[5]. The argument in favor of insufficient capacity is the reverse of the one used for excess capacity: with insufficient

---

[1]Phlips (1995*a*) pages 151–153

[2]In terms of firms' participation in the cartel.

[3]In the sense of being unable unable to sustain a high price.

[4]Another common definition of capacity, used in the monopolistic competition literature, is the level of production that minimizes the average cost (Phlips 1995*b*).

[5]If the marginal revenue is positive.

capacity firms cannot harshly punish deviations from the tacit agreement, but then deviations are not as profitable either. Therefore, it would be only reasonable that a model used to study excess capacity should allow for the possibility of insufficient capacity, too. I construct such a model in this paper. A firm's production function is defined over two inputs: capital and labour. The latter is deemed as a variable input whereas the former is chosen prior to production and remains fixed subsequently. I interpret the capital input as capacity[6], and the optimal capacity as the level of capital that minimizes the total cost of producing a given level of output.

Firms can choose to carry excess capacity for various reasons: to accommodate peak-demands; as a response to rate-of-return regulation (blamed for the overcapacity of British railroads e.g.) or some other type of regulation (regulation allocating output/profits on basis of capacity is blamed for over-drilling for oil in Texas in 1930s (Davidson 1963)); to strategically deter entry, as in the model of Dixit (1980); to increase its power in explicit or implicit bargaining with other firms, as in the analysis of Osborne & Pitchik (1987); possibly to help tacit collusion, by increasing punishment capabilities.

My paper investigates only whether it is excess capacity or insufficient capacity that helps tacit collusion, and I am going to argue that the usual result of tacit collusion with excess capacity is due to firms' use of capacities for rent-seeking purposes.

Brock & Scheinkman (1985) is one of the first papers to challenge the traditional view, demonstrating that a certain amount of excess capacity is required to enforce a cartel. Unfortunately their model assumes exogenous capacity and, therefore, while in their model increasing capacity may help collusion in prices in a supergame, they cannot conclude that excess capacity emerges in equilibrium. In this sense, their model analyzes the effect of capacity constraints in enforcing collusion in a supergame, not the role played by excess capacity in tacit collusion. They conjecture that if the capacity is endogenous and the resulting equilibrium involves collusion at the monopoly price, then there should be excess capacity.

---

[6]Similar to Eaton & Eswaran (1997)

Benoit & Krishna (1987) endogenize the choice of capacity by the members of a cartel in a model in which firms first choose the capacity, knowing that they are subsequently going to play an infinitely repeated price game. The main result of their analysis is that, in all[7] subgame perfect equilibria of the model, firms carry excess capacity. Note that in their model firms (tacitly) collude in both prices and capacities, and collusion is enforced by 'double-barrelled' trigger strategies that punish deviations in capacity or price. Benoit & Krishna (1987) also show that their result is robust with respect to capacity adjustments in every period, provided firms can undertake additional investment in capacities only in small steps.

Davidson & Deneckere (1990) consider a particular class of equilibria of a model very similar to that used by Benoit & Krishna (1987): firms tacitly collude in prices but not in capacities. That is, for given capacity choices, firms charge the highest sustainable price that can be supported by a collusive agreement. Any deviation from this pricing is punished by permanently reverting to the one-period Nash equilibrium prices. Since the equilibria of this model are contained in the set of equilibria analyzed by Benoit & Krishna (1987), excess capacity emerges in all subgame perfect equilibria. They find that increases in the interest rate and/or decreases in capacity costs lead to higher levels of collusion and capacity.

Osborne & Pitchik (1987) derive a result of excess capacity in a model in which firms first choose capacity non-cooperatively, knowing that their choice is followed by explicit collusion in prices, with the final profit being allocated via Nash bargaining. The excess capacity is due to the fact that capacity improves a firm's bargaining power and, therefore, each firm races to install excess capacity in the hope of obtaining a higher profit share.

The above analyses might create the impression that excess capacity helps collusion. But the common feature of these studies is not that firms face capacity constraints, but rather that the firms with larger capacities in the second stage of the game (supergame in prices or explicit collusion with Nash bargaining) get higher shares of the total profit.

---

[7]Except those equilibria that mimic a Cournot-Nash equilibrium.

This is the force that drives the excess capacity result, and not the fact that excess capacity helps collusion. The race to get a higher share of the profit can be avoided in two ways. One is to allow firms to collude explicitly[8] in capacities, a case in which Eaton & Eswaran (1997) show that firms are likely to choose insufficient capacity. Alternatively, one could have firms invoke a type of tacit agreement that does not allocate the profit according to the capacity of each firm. My analysis follows this route, and demonstrates that insufficient capacity can emerge in equilibrium.

My analysis considers a two-stage game similar to that used by Davidson & Deneckere (1990): in the first stage, firms non-cooperatively and irreversibly choose capacities, and the second stage there is a supergame in quantities[9]. The tacit contract is different from that used by Davidson & Deneckere (1990) in the sense that (ex-ante) identical firms tacitly agree to produce the same level of output, regardless of their ex-post capacity levels. Note that this is not an explicit collusion contract; this agreement has to be self-enforcing and, therefore, it is subject to the usual incentive constraints. For example, the common output level cannot be very low, because the competitor would find it profitable to deviate by producing more. I believe this is a very reasonable agreement for firms that are *ex ante identical.* They can easily figure out that in equilibrium they will have equal capacities, almost regardless of the type of tacit agreement they are going to use. The adoption of an egalitarian agreement, which avoids the perils of a capacity race, is potentially a profitable decision.

The major advantage of this type of agreement (referred to from now on as the 'egalitarian agreement') is its simplicity: once the output is agreed on, checking for deviations is very easy, involving just a quantity estimation. Another advantage is that avoids the standard rent-seeking outcome that leads firms to overinvest in capacities to capture a higher profit share. My point of departure is also reasonable since rent-seeking

[8]The assumption of explicit collusion, likely to be forbidden by most antitrust laws, reduces the attractiveness of this avenue of research.

[9]Since this is a simultaneous moves game, I believe the assumption of quantity competition supergame is not crucial and it would yield similar results (Brander & Harris (1984) show that excess capacity can also emerge in a quantity competition model).

behaviour hurts firms and gives them an incentive to search for alternative types of tacit agreements[10]. With a large discount factor (but strictly less than 1) and an egalitarian agreement, industry profit can reach the monopoly level. But, because of the excess capacity, it is unable to achieve monopoly profit with the usual agreement used by Davidson & Deneckere (1990) even when the cartel is able to sustain the monopoly price.

As we shall see, the main drawback of the egalitarian agreement is that it cannot be enforced for low values of the discount factor. The tacit agreement usually invoked in the literature specifies that firms charge the highest price supported by the capacities they already have. In other words, the usual contract specifies that each firm charges a price depending on its capacity, and checking for deviations calls for an estimate of rivals' capacities, too. Furthermore, the relationship between price/quantity and capacity might be quite simple in a Leontieff world, but becomes extremely complicated even with the simple Cobb-Douglas production function. Even though my model does not include imperfect information considerations, I believe that the egalitarian agreement would be cheaper to implement in a world in which a low price could be caused either by unobserved demand shocks or by deviant firms, like in the study of Green & Porter (1984).

If firms use the tacit egalitarian agreement, I demonstrate that there is an incentive to reduce the capacity below the cost minimizing level: reducing capacity signals to a firm's rivals that it does not intend to deviate and therefore gives incentives to the rivals to agree to a lower common output. Reducing capacity is a way to commit to collusion. The more costly is capacity, the less is the demand for capacity as an instrument facilitating collusion and, therefore, the installed capacity tends to *increase* relative to the optimal capacity when capacity cost increases.

[10]Brander & Harris (1984) show, in a model with explicit collusion in the second stage, that firms might be worse off than without collusion if the profit share each firm gets is equal to its share of industry capacity.

Since even the simple model with Cobb-Douglas technologies does not lend itself to analytic solutions, I have to use numeric methods to explore the equilibria of my model. Nevertheless, I isolate the economic intuition driving the results and I identify when the results are expected to hold.

I find that increased collusion due to an exogenous increase in the discount factor can be welfare increasing. The increased discount factor, by facilitating collusion, reduces the pressure to curtail capacity and, therefore, firms become more cost-efficient. This gain in efficiency may be greater than the decrease in consumer surplus due to higher prices. Increased collusion is accompanied by increased capacity, but the capacity still remains lower than the cost-minimizing level, until full collusion is reached (at high discount factors).

I also find that welfare is *not monotonic* with the number of firms. As the number of firms grows, the increased competition decreases the profitability of cheating and firms can agree on a slightly smaller output, leading to a welfare decrease. Ultimately, as the number of firms increases, the egalitarian agreement cannot be enforced and the addition of firms becomes welfare enhancing. It must be emphasized that all these results obtain in a model in which the production technology exhibits constant returns to scale, and so there are no built-in effects arising from size.

The model has several antitrust policy implications. First, in some circumstances, insufficient capacity can be an indicator of tacit collusion. Second, a policy that promotes entry does not necessarily increase welfare, even in the absence of fixed costs. Such a policy is guaranteed to increase welfare only if it manages to attract a sufficiently large number of entrants. Thus, reliance on indices like concentration ratios, Herfindahl index, etc. can lead to erroneous conclusions regarding the performance of industries even when all firms are symmetric.

2. The Market, Technology and Capacity

In this section, I set out a model that is equipped to address the issues raised in the Introduction.

2.1. The market. Firms compete in quantities in a market characterized by the inverse demand function:

$$p = 1 - Q \tag{2.1}$$

where Q is the total amount of output sold in the market.

I am assuming there is no entry in this market.

2.2. Technology. All firms have identical Cobb-Douglas technologies:

$$q(k, l) = k^{1/2} l^{1/2} \tag{2.2}$$

where k, l are the amount of capital and labor employed for producing the output q.

The above production function yields a restricted cost function, the cost of producing q units of output when capacity is fixed at k:

$$C(q, k) = rk + w\frac{q^2}{k} \tag{2.3}$$

where q is the output and r, w are the rental rate of capital and labor, respectively.

I assume that capital is less flexible than labour and I am loosely interpreting it as capacity. Capital is the first input chosen and will not change thereafter, while labour can change every period.

Given an output q, I define the optimal capacity, k^C, as the cost-minimizing capacity. Mathematically, k^C solves the equation:

$$\frac{\partial C(q, k^C)}{\partial k} = r - w\frac{q^2}{k^{C2}} = 0 \tag{2.4}$$

Since the cost function is convex, a value of capacity for which $\frac{\partial C(q,k)}{\partial k} < 0$ means that there is 'insufficient capacity' ($k < k^C$), whereas if $\frac{\partial C(q,k)}{\partial k} > 0$ there is 'excess capacity' ($k > k^C$).

3. Two firms case

There are two firms in the market and they compete in two stages. In the first stage they irrevocably choose and build capacities. In the second stage they play an infinitely repeated game in quantities.

I solve for the subgame perfect equilibria by the standard procedure of backward induction: first I solve for the subgame perfect equilibria in the second stage (the supergame in quantities) for arbitrary initial capacity choices[11],then I step back to the first stage and work out the equilibrium capacity choices.

3.1. Second stage. The instantaneous profit of firm i, given its capacity choice, is:

$$\pi_i(q_i, k_i, q_j) = (1 - q_i - q_j)q_i - rk_i - w\frac{q_i^2}{k_i} \tag{3.1}$$

where $i, j = 1, 2$ and $i \neq j$.

The best reply function of firm i is easily seen to be:

$$R_i(q_j) = \frac{k_i\ (1 - q_j)}{2\ (w + k_i)} \tag{3.2}$$

Firm i's best response to a given output of its rival is increasing in its own installed capacity.

The static Cournot-Nash equilibrium output of firm i is readily shown to be:

$$q_i^{SNE}(k_i, k_j) = \frac{k_i\ (2\,w + k_j)}{4\,w^2 + 4\,w\,(k_i + k_j) + 3\,k_i k_j} \tag{3.3}$$

and the corresponding profit to be:

$$\pi_i^{SNE}(k_i, k_j) = \frac{k_i\ (w + k_i)\ (2\,w + k_j)^2}{\left(4\,w^2 + 4\,w\,(k_i + k_j) + 3\,k_i k_j\right)^2} - rk_i \tag{3.4}$$

Firms tacitly collude using a Nash reversion strategy[12]. Firms tacitly agree to a collusive output, threatening that any deviation will be punished by forever producing the one period Nash equilibrium output. There are many ways in which the collusive output could be set. Since a firm's ability to punish increases with its capacity (because of lower marginal cost), perhaps the most natural tacit agreement would be to let the firm with the larger capacity to produce more. But this is not the only possible self-enforcing

[11]And since I am interested only in symmetric equilibria, I assume firms do not choose very different capacities. The subgame perfect equilibrium is checked to see whether it is robust to large capacity deviations in the first stage.

[12]For analytical tractability, I do not consider optimal punishment strategies, but there is no compelling reason to believe they would change the qualitative results of the model.

agreement. I believe it is also quite reasonable that firms agree to produce the same output, regardless of installed capacity, especially for firms that are identical ex-ante. They already know that in equilibrium they will end up producing the same output, so it is also natural that they agree ex-ante to produce the same amount of output, avoiding the losses due to rent-seeking behavior. Furthermore, this egalitarian agreement is simpler and therefore easier to check for deviations in an imperfect information world[13], since it requires a firm to check *only* whether the other firm produced the agreed output.

Therefore, the grim trigger strategy I consider assumes firm i produces the same output as firm j, denoted by q_a, as long as both firms always produced q_a in the past; deviations are punished by reverting permanently to the static Nash-Cournot level $q_i^{SNE}(k_i, k_j)$.

The usual tacit agreement used by Davidson & Deneckere (1990) allows the agreed-upon output for firm i to depend on its installed capacity k_i, thereby possibly generating non symmetric equilibria in the quantity supergame. The advantage of this kind of tacit collusion is that it is enforceable for any value of the discount rate ($\delta \in (0, 1]$) and for arbitrary capacities. But in practice deviation from tacit agreement is much harder to detect, since it requires estimation of both output and capacity. In contrast, as we will see, the egalitarian agreement employed here cannot be enforced for all capacity differences between competitors or for low values of the discount factor, but it is much simpler to detect deviations, making it easier for firms to collude.

Another type of trigger strategies used in the literature[14] is the so called 'double barreled' strategies, in which the trigger depends on two strategic variables: in our case, the trigger would depend on both output and capacity, and would allow tacit collusion in both capacity and output. I do not allow any collusion in capacities, however, on the ground that empirical evidence suggests that firms find difficult to collude on long-term strategic decisions.

---

[13]That is, a world in which rival's output and capacity can be imprecisely observed, not a world in which they cannot be observed at all (as in Green & Porter (1984)).

[14]For example, by Benoit & Krishna (1987) and Eaton & Eswaran (1997).

The one-period profit, $\pi_i^C(q_a, k_i)$, firm i derives from cooperation in stage 2 is:

$$\pi_i^C(q_a, k_i) = (1 - 2\,q_a)\; q_a - \frac{w\,{q_a}^2}{k_i} - r\,k_i \tag{3.5}$$

By deviating optimally, the deviation profit, $\pi_i^D(q_a, k_i)$, of firm i is:

$$\pi_i^D(q_a, k_i) = \pi_i(R_i(q_a), k_i, q_a) = \frac{k_i\,(1 - q_a)^2}{4\;(w + k_i)} - r\,k_i \tag{3.6}$$

For the egalitarian agreement to be self-enforcing, we require that the present value profit from cooperation be at least as high as that from deviation. Mathematically, this no deviation condition is:

$$\frac{\pi_i^C(q_a, k_i)}{1 - \delta} \geq \pi_i^D(q_a, k_i) + \frac{\delta}{1 - \delta}\pi_i^{SNE}(k_i, k_j) \quad , \tag{3.7}$$

where δ is the discount factor, assumed to be common between the two firms. In current period terms, this can be rewritten as:

$$\pi_i^C(q_a, k_i) - (1 - \delta)\pi_i^D(q_a, k_i) - \delta\pi_i^{SNE}(k_i, k_j) \geq 0 \quad . \tag{3.8}$$

The left hand side of equation (3.8) is a concave quadratic function in q_a. Labelling its roots by $\underline{q}_a(k_i, k_j)$ and $\overline{q}_a(k_i, k_j)$ (such that $\underline{q}_a(k_i, k_j) \leq \overline{q}_a(k_i, k_j)$), we know that this kind of agreement can credibly induce firm i to produce any output $q_i \in [\underline{q}_a(k_i, k_j), \overline{q}_a(k_i, k_j)]$.

Invoking symmetry, firm j can be convinced to produce any output $q_j \in [\underline{q}_a(k_j, k_i), \overline{q}_a(k_j, k_i)]$.

Condition (3.8) has solutions for q_a when $k_i \leq k_j$, since firm j has superior punishing ability. It cannot be satisfied for any output q_a when k_j is much lower than k_i, when the low capacity renders firm j incapable of influencing firm i decision in any way.

For the moment I am limiting my analysis to the case where k_i is not too different from k_j. It can be proven that in the case $k_i = k_j = k > 0$, the limits $\underline{q}_a(k, k)$ and $\overline{q}_a(k, k)$ always exist and that they are positive. By continuity, it follows that there is a vicinity of $(k_i, k_j) = (k, k)$ for which the no deviation condition (3.8) can hold for both firms. Without loss of generality, from now on I consider only the case $k_i \geq k_j$, with k_i close to k_j. The other case can be considered by simply swapping the two firms.

Label by $NDC_i(k_i, k_j)$ the left hand side of the inequality (3.8) for firm i. By symmetry, $NDC_j(k_i, k_j) = NDC_i(k_j, k_i)$. It is easy to verify that:

$$NDC_i(k_i, k_j) - NDC_j(k_i, k_j) =$$

$$(k_j - k_i)\left[r + (1-\delta)\frac{w(1-q_a)^2}{4(w+k_i)(w+k_j)} + \delta\frac{(k_i+k_j+w)(k_j+2w)^2}{4w^2+4w(k_i+k_j)+3k_ik_j}\right] \tag{3.9}$$

If NDC_i is positive (i.e. the no defection condition is satisfied for Firm i), then, from above equation, $NDC_j \geq NDC_i$ and therefore NDC_j is positive too. In other words, if the high capacity firm would produce an output q_a without deviating then so would the low capacity firm, but the reverse is not necessarily true. This proves that $\underline{q}_a(k_i, k_j) \geq \underline{q}_a(k_j, k_i)$ and $\overline{q}_a(k_i, k_j) \leq \overline{q}_a(k_j, k_i)$ for $k_i \geq k_j$. Intuitively, the firm with higher installed capacity finds easier to deviate from the agreement, and therefore the range of output it can be induced to produce is *smaller* than that of its competitor. Since firm i can cooperate by only producing an output within the interval $q_i \in [\underline{q}_a(k_i, k_j), \overline{q}_a(k_i, k_j)]$ and firm j only by producing within the interval $q_j \in [\underline{q}_a(k_j, k_i), \overline{q}_a(k_j, k_i)]$, they will both agree to a common level of output $q_i = q_j = q_a$ if they produce an output within the intersection of these two intervals. That is, this kind of egalitarian trigger strategy will support any output, q_a, within a certain interval which I write as:

$$q_a \in \left[\underline{q}_a(k_i, k_j), \overline{q}_a(k_i, k_j)\right] \tag{3.10}$$

Finally, the two firms in the second stage maximize the joint profit by choosing the common level of output that is compatible with the no deviation conditions of the two firms. Formally, the problem the two firm faced in the second stage of the game is:

$$\max_{q_a} \quad \pi_1(q_a, k_1, q_a) + \pi_2(q_a, k_2, q_a) \tag{3.11}$$

subject to the no deviation condition (3.8) for $i = 1, 2$.

While the joint profit maximization problem (3.11) might seem puzzling in absence of side-payments, remember that symmetry dictates that both firms install the same capacity in equilibrium, and they both obtain half of the cartel's profit by using the

egalitarian agreement. Therefore, by maximizing its profit, the cartel maximizes the profit of each of its members.

Since the objective function in (3.11) is concave in q_a, the solution is given by:

$$q_a(k_i, k_j) = \max \left\{ \underline{q}_a(k_i, k_j), \arg\max_q \left[\pi_i(q, k_i, q) + \pi_j(q, k_j, q) \right] \right\} \quad . \tag{3.12}$$

In other words, if the unconstrained optimum of joint profits satisfies the no defection conditions, then that is the solution. If not, the solution is the lowest output that satisfies these. I assume[15] that $\overline{q}_a(k_i, k_j)$ is always higher than the unconstrained (monopoly) solution to problem (3.11).

3.2. First stage. For the moment, I consider only the case where the discount factor is sufficiently low such that the constrained solution to (3.11) is always higher than the monopoly solution. This is also the more interesting scenario. The second stage's solution, then, is:

$$q_a(k_i, k_j) = \underline{q}_a(k_i, k_j) \tag{3.13}$$

By cooperating in the second stage, firm i realizes a profit $\Pi_i(k_i, k_j)$

$$\Pi_i(k_i, k_j) \equiv \pi_i \left(\underline{q}_a(k_i, k_j), k_i, \underline{q}_a(k_i, k_j) \right) \tag{3.14}$$

In the first stage, in which the capacity is non-cooperatively determined, under the Nash conjectures the firm solves the problem:

$$\max_{k_i} \quad \Pi_i(k_i, k_j) \tag{3.15}$$

The first order condition of problem (3.15) is:

$$\frac{\partial \Pi_i(k_i, k_j)}{\partial k_i} = 0 \tag{3.16}$$

Invoking symmetry, we have $k_i = k_j = k^*$ where k^* solves

$$\frac{\partial \Pi_i(k^*, k^*)}{\partial k_i} = 0 \tag{3.17}$$

---

[15] In what follows, I check that this assumption is indeed satisfied in equilibrium.

Although equation (3.17) does not yield an analytic solution, it can be solved numerically.

Note that equation (3.17) is a necessary condition for a local maximum only. It is possible that given $k_j = k^*$, firm i chooses $k_i = k^*$ if we restrict the analysis to a close neighborhood of k^*. If we extend the domain to the range $k_i \in (0, \infty)$, firm i may want to increase its capacity beyond the critical level that makes tacit collusion possible, preferring instead the Cournot equilibrium in which it has a much higher capacity than its rival. This happens for very low discount factors, when basically any tacit agreement brings a very low increase in profit relative to the Cournot equilibrium.

Figure 13 depicts this possibility. Firm j has a capacity k^*. (The vertical axis meets the horizontal one at $k_i = k^*$.) The thick curve graphs the profit of firm i, $\Pi_i(k_i, k^*)$, as a function of its own capacity. Note that the curve ends suddenly because when k_i is much higher that k^* collusion cannot be enforced. Firm i can also choose to never cooperate in the supergame, obtaining a profit $\pi^{SNE}(k_i, k^*)$ and this is graphed by the thin curve (which reaches its maximum for capacity values so much higher than k^* that it does not fit into the picture). As we can see, if firm i decides to cooperate, it will choose a capacity $k_i = k^*$, where the $\Pi_i(k_i, k^*)$ reaches its maximum. If it decides not to cooperate, it will choose $k_i \gg k^*$ where the $\pi^{SNE}(k_i, k^*)$ reaches its maximum (not presented in figure). Since the Nash maximum profit is higher than the collusion equilibrium, firm i does not cooperate and therefore (k^*, k^*) is not a part of the subgame perfect equilibrium if the discount factor is very low. In other words, the tacit collusion profit line offers a local maximum, but a firm will prefer (in this case) to deviate from the local maximum and so the firms do not collude in the supergame.

An increase in the value of the discount factor shifts the $\Pi_i(k_i, k^*)$ line up and rightward so that it eventually has a higher maximum than $\pi^{SNE}(k_i, k^*)$ and collusion becomes enforceable, with both firms choosing a capacity of k^* in the first stage of the game. If the discount factor increases even more, ultimately, the solution to (3.11) is the monopoly solution, and the constraints do not bind. Firms do not need to restrict

capacity to sustain collusion and therefore they choose the cost-minimizing level of capacity.

There are two critical values, δ_N and δ_M, of the discount factor, with $0 < \delta_N < \delta_M < 1$ and such that:

- If $\delta < \delta_N$, then the second stage equilibrium is the static Nash equilibrium and involves no tacit collusion;
- If $\delta_N \leq \delta \leq \delta_M$, both firms tacitly collude to produce some output q_a which depends on δ and is more than half the monopoly output in each period of the second stage;
- If $\delta_M < \delta$, firms achieve perfect collusion in quantities, that is, they produce the monopoly output.

The main result of this paper, the claim that insufficient (relative to the cost-minimizing level) capacity facilitates tacit collusion, is confirmed[16] by the simulations to follow. The insufficient capacity is caused by the strategic role played by capacity. By decreasing capacity, a firm reduces its deviation profit, making it easier for the rival to agree to a lower collusive output. The rival has less reason to be wary of a firm that has curtailed its ability to deviate from the tacit agreement. Therefore, a firm has an incentive to decrease its capacity below the cost minimizing level, in order to facilitate collusion.

3.3. Comparative statics. For the moment I restrict the comparative statics to the case $\delta \in [\delta_N, \delta_M]$, where tacit collusion with an egalitarian agreement is feasible.

Figure 14 graphs the relationship between the optimal capacity and rental rate, r, of capacity. Recall that k^* denotes the (common) equilibrium capacity level that maximizes their feasible joint profits, while k^C denotes the capacity level that minimizes the cost of producing each firm's output in this equilibrium. Since the amount of input used varies inversely with input price, both k^* and k^C monotonically decrease with r. The difference

[16] To check the cost-minimizing level of capital we have to evaluate equation (2.4) for $q = \underline{q}_a(k^*, k^*)$ and $k = k^*$.

$k^C - k^*$ is also monotonically decreasing with r. When the rental rate declines, the cost-minimizing capacity naturally increases; however, the equilibrium capacity increases by less because low capacity levels facilitate tacit collusion.

Because labor and capacity are substitutes in production, increases in wage rate have exactly the opposite effect (See Figure 15). Small increases (from a low level) in wage increase the input demand for capacity and therefore both k^* and k^C increase with w. Since the capacity increase hampers tacit collusion, firms increasingly sacrifice production efficiency to facilitate collusion and the difference $k^C - k^*$ increases with w.

Figure 16 graphs the effect of the discount factor on equilibrium capacity. With a low discount factor ($\delta < \delta_N$), collusion cannot be sustained and firms produce the Cournot equilibrium level of output: the optimal capacity is $k^* = k^{SNE}$. With high discount ($\delta \geq \delta_M$), factor firms achieve monopoly results in the supergame and the optimal capacity is $k^* = k^M$. Since they do not need to restrict capacity to aid collusion, the strategic role played by capacity disappears and firms produce 'at capacity'.

When the discount factor is in the intermediate range $[\delta_N, \delta_M)$, firms tacitly collude but cannot replicate the monopoly result. Since an increase in the discount factor makes deviation less profitable, firms can afford higher levels of capacity. Since the purpose of collusion is to decrease the output in the market, the capital required to minimize cost for that output decreases too. Because of the greater feasibility of collusion brought about by the increased discount factor, firms have less incentive to sacrifice technological efficiency to facilitate collusion. Thus while k^C decreases with the discount factor, k^* increases.

Figure 17 depicts the dependence of the equilibrium output, denoted q^*, on the discount factor. For a low discount factor collusion cannot be enforced and therefore the output in the subgame perfect Nash equilibrium is constant. When the discount factor becomes larger than the critical value δ_N, collusion becomes possible and the output jumps down discontinuously. As the discount factor keeps increasing, the tacit collusion approaches the monopoly equilibrium, which is reached for $\delta = \delta_M$. If the

discount factor continue to increase, the optimal output, of course, remains constant at the monopoly level.

Note that with a discount factor increasing in the range $[\delta_N, \delta_M)$, q^* and $k*$ move in *opposite* directions: q^* decreases because firms have less incentive to deviate from agreed-upon output levels. The reduced incentive to deviate allows firms to be more cost efficient by increasing their capacity.

As soon as the discount factor becomes larger than δ_N and firms can enforce tacit collusion, their profit jumps up, as it is shown in Figure 18. As the discount factor increases, the degree of collusion and profits increase too, until the discount factor reaches the next threshold level, δ_M, point at which profits reach the maximum level. The discontinuous increase in profits at $\delta = \delta_N$ is easily explained by Figure 13. Recall that tacit collusion requires that the thin (Nash) profit line has a maximum below the maximum of the thick (tacit) profit line. In particular, since the tacit profit line reaches its maximum at the intercept with the vertical axis, the Nash profit has to intersect the vertical axis at a point strictly below the tacit profit line. Note that this discontinuous profit increase at $\delta = \delta_N$ is due mainly to the fact that the duopoly manages to restrict its output, and not only because the firms become more technologically efficient. In the Nash-Cournot equilibrium, firms over-invest in capacity, while in the semi-collusive equilibrium firms are still not efficient because they under-invest in capacity.

As Figure 19 shows how the welfare generated by the industry per period in stage 2 varies with the discount factor. The measure of welfare adopted here is the standard one of total surplus (that is, consumers' surplus plus producers' surplus). As can be seen from the Figure, welfare decreases discretely when the firms switch from a Nash to a collusive equilibrium for a reason that should be familiar by now: firms curtail their output, while remaining technologically inefficient. As δ increases, firms continue to further curtail their output. One might expect welfare to be non-increasing in the discount factor, since higher δ facilitates greater collusion. Notice, however, that the dependence of welfare on the discount factor is *non-monotonic* in Figure 19. When δ is slightly higher than δ_N, the main benefit of collusion consists in allowing firms to

increase their capacities, accompanied by a slight decrease in output. The increased production efficiency more than compensates for the decrease in welfare due to lower output, and generates an increase in total surplus over this range. As capacities continue to grow, their marginal benefit decreases and firms find it more and more profitable to curtail output. Ultimately, the decrease in consumer surplus overwhelms the benefit of cost-reduction, and tacit collusion decreases welfare.

The comparative statics of welfare with respect to input prices are routine, as long as the firms do not switch from a semi-collusive equilibrium to a non-collusive equilibrium. In this case welfare monotonically decreases with both input prices.

Changes in input prices, however, can induce an equilibrium switch by changing the value of δ_N. That is, input prices affect the range of values of the discount factor for which the tacit collusion is possible. Simulations indicate that δ_N increases monotonically with both r and w. If the actual discount factor of the firms is near but greater than the critical value δ_N, an increase of sufficient magnitude in at least one of the input prices can raise the value of δ_N, making the collusion unenforceable. The industry then switches from an equilibrium entailing tacit collusion to a Cournot-Nash equilibrium, and, since such a switch is accompanied by a discrete increase in welfare, an increase in input prices can *increase* the total surplus. In Figure 20, as a result of an increase in the wage rate, the firms move from point A (where firms collude) to point B (where collusion cannot be supported), thereby increasing welfare. Thus, welfare can be non-monotonic in input prices.

4. n-firm case

In a market described by the same inverse demand function (2.1), I now assume that there are n firms having access to the technology (2.2). The interpretation of the optimal capacity level is the same as before. The subgame perfect equilibrium is computed by backward induction.

4.1. Second stage. The instantaneous profit of firm i is:

$$\pi_i(q_i, k_i, Q_{-i}) = (1 - q_i - Q_{-i})q_i - rk_i - w\frac{q_i^2}{k_i} \tag{4.1}$$

where $i = 1 \ldots n$ and Q_{-i} is the total output produced by all firms other than i ($Q_{-i} = \sum_{j \neq i} q_j$).

The best reply function of firm i is easily seen to be:

$$R_i(Q_{-i}) = \frac{k_i\ (1 - Q_{-i})}{2\ (w + k_i)} \tag{4.2}$$

I am looking for symmetric equilibria, and therefore I will assume that while firm i has a capacity k_i, each of the other firms has the same capacity k and produces the same output q. The reaction function of firm i is given above, where $Q_{-i} = (n-1)q$, and the reaction function of one of the other firms is:

$$R_j(Q_{-j}) = \frac{k\ (1 - Q_{-j})}{2\ (w + k)} = \frac{k\ [1 - (q_i + (n-2)q)]}{2\ (w + k)} \tag{4.3}$$

where $j \neq i$.

The static Cournot-Nash equilibrium is easily found by simultaneously solving $q_i = R_i(Q_{-i})$ and $q = R_j(Q_{-j})$. The corresponding output and profit are denoted by $q_i^{SNE}(k_i, k, n)$ and $\pi_i^{SNE}(k_i, k, n)$.

Firms tacitly collude on an egalitarian agreement: firm i produces the same output as all the other firms, q_a, as long as all firms produced q_a in the past; deviations trigger a permanent play of the Cournot-Nash equilibrium output, $q_i^{SNE}(k_i, k, n)$. As before, the tacit agreement is subject to incentive constraints of the type (3.8).

The procedure for finding q_a is very similar to the two firms case. In the second stage, firms choose a common level of output to maximize the joint profit. Formally, the problem is:

$$\max_{q_a} \quad \pi_i(q_a, k_i, q_a) + (n-1)\pi_j(q_a, k, q_a), \qquad j \neq i \tag{4.4}$$

subject to the no deviation condition (3.8). The solution to this problem is denoted by $q_a(k_i, k, n)$.

Note that, as before, q_a does not exist if k_i is very different from k. Without loss of generality one can assume that $k_i \geq k$.

4.2. First stage. The procedure used in the two firm case is easily adapted for the n firm case. Cooperating in the second stage, firm i realizes a profit $\Pi_i(k_i, k, n)$:

$$\Pi_i(k_i, k, n) \equiv \pi_i \left(q_a(k_i, k_j), k_i, (n-1)q_a(k_i, k_j)\right) \tag{4.5}$$

In the first stage, in which the capacity is non-cooperatively determined, the firm solves the problem:

$$\max_{k_i} \quad \Pi_i(k_i, k, n) \tag{4.6}$$

The first order condition of problem (4.6) is:

$$\frac{\partial \Pi_i(k_i, k, n)}{\partial k_i} = 0 \tag{4.7}$$

Invoking symmetry and looking for a symmetrical equilibrium, we have $k_i = k = k^*$ where k^* solves

$$\frac{\partial \Pi_i(k^*, k^*, n)}{\partial k_i} = 0 \tag{4.8}$$

Once again, equation (4.8) does not yield to an analytic solution, but it can be solved numerically.

Like in the two firm case, for every n, there are two critical values, δ_N and δ_M, of the discount factor. For $\delta < \delta_N$ firms cannot sustain any collusion in the supergame, while they are able to tacitly collude for $\delta \in [\delta_N, \delta_M)$, and replicate the monopoly outcome if $\delta \geq \delta_M$.

4.3. Comparative statics. Simulations indicate that both δ_N and δ_M monotonically increase with the number of firms, n, while the difference $\delta_M - \delta_N$ monotonically decreases with n. In other words, the range of discount factors that can support partial and full collusion decreases with n. The intuition is simple: an increase in the number of competitors hinders collusion by increasing the incentive to cheat.

As expected, the capacity of each firm monotonically decreases with n, no matter how well firms are able to collude.

In the case in which firms are able to replicate the monopoly result in the supergame, industry output, as a function of n, remains constant and firms are cost-efficient. This is due to the fact that the tacit egalitarian agreement I am considering does not offer any incentive for a firm to increase its capacity above the optimal level. Once firms in perfect collusion agree to produce a common level of output, the only role of the capacity is to minimize costs, not to facilitate collusion. There is a linear relationship between output and optimal capacity, given the linearly homogeneous technology. Thus the aggregate industry capacity does not depend on n if the firms can perfectly collude to produce the monopoly output. An increase in the number of firms naturally results in a smaller market share for all firms, and each firm uses the optimal capacity to produce its share. This case is illustrated in Figure 21 for $n = 1, \ldots, 4$. This contrasts with the results of Benoit & Krishna (1987) and Davidson & Deneckere (1990): in their model, even when firms are able to reproduce the monopoly outcome in prices, the industry earns a lower profit than the monopoly profit because they have excess capacity.

In the case where the firms cannot collude at all (that is, for low discount factors or for large numbers of firms), the profits in the supergame are distributed according to the capacity chosen in the first stage. The firm with a lower marginal costs (that is, higher capacity) gets a higher share of the collusive output and, therefore, obtains a higher profit. This strategic effect is the reason that, in the subgame perfect equilibrium, firms have incentives to increase their capacity above the efficient level. These incentives are greater when the number of firms is relatively small and, therefore, each firm's capacity decreases slowly and industry capacity increases with n when n is large enough that no collusion is possible, but small enough that the strategic effect is not negligible. This is illustrated in Figure 21, for $n \geq 8$.

Incentives to use excess capacity greatly decreases with large n since competition becomes more intense and the diminishing market share dilutes the incentive to install capacity for strategic purposes. Therefore, industry capacity decreases for very large n. This case is presented in Figure 22.

In the intermediate case where firms manage to partially collude ($4 < n < 8$ in Figure 21), capacity is decreasing with the number of firms. Since an increase in n tends to hinder collusion by encouraging cheating, firms need to decrease capacity to discourage cheating. The capacity decrease per firm is so great, in fact, that industry capacity decreases even as n increases.

Note that even though industry capacity decreases with n for both the intermediate and for very large values of n, the decrease is due to different causes. In the intermediate range of n, where there is imperfect collusion, firms decrease capacity to facilitate collusion (and all firms install less capacity than the cost-minimizing one). In the case with a very large number of firms, no tacit collusion based on an egalitarian agreement is possible and the decline in capacity is due to the fact that capacity is less profitable in the strategic role of 'stealing' market share from the other firms, and so aggregate capacity declines in n. Nevertheless, here all firms still have some excess capacity.

As we have already seen, when firms are able to collude fully, industry capacity and output are independent of n. If the number of firms is large, industry output increases with n for the standard reason: the increased competition diminishes the underproduction associated with imperfect competition. Note that industry output increases even when aggregate capacity declines. Compare Figures 21 and 23 over the range $4 \leq n \leq 8$. In this range, an increase in n increases the aggregate equilibrium output, but this comes largely from a more than commensurate increase in the employment of the variable factor, labour. The aggregate decline in capacity results from the attempt to reduce incentives to cheat, as we have already seen.

Figure 23 shows the dependence of industry output on the number of firms. When firms partially collude, each firm's output declines with n but industry output can vary non-monotonically with n: it can increase with n for small n and decrease with n for large n. As the number of firms grows, firms may find it easier to partially collude, in the sense that cheating becomes less profitable (similar to the way in which industry capacity declines with n when there is a large number of firms that are not colluding). Therefore, firms find it easier to agree to smaller individual outputs, and it can happen

that the individual output declines faster than $1/n$, leading to a decline in the industry output. It is somewhat difficult to see this in Figure 23, but careful scrutiny reveals that industry output declines when the number of firms increases from 6 to 7.

Figure 24 presents the effect of the number of firms in the market on welfare. With only a few firms and high discount factor, firms perfectly collude and, since the production function exhibits constant returns to scale, welfare does not depend on n. When n increases sufficiently, firms cannot perfectly collude any longer and they use deficient capacity to enforce collusion. Output and welfare then increase with n. If n continues to increase, firms have more incentives to restrict their capacity, leading to greater cost-inefficiency and this can lead to a *decrease* in welfare. This welfare decrease due to the cost-inefficiency can be reinforced by a decline in consumer surplus caused by a declining aggregate output, as discussed above. In Figure 24, this decline in welfare can be seen when n increases from 6 to 7. As n continues to increase, ultimately firms cannot collude at all; output and welfare then increase.

In summary, I have demonstrated that, despite the fact that firms here are using linearly homogenous technologies, with an egalitarian tacit agreement:

(1) If firms are able to replicate the monopoly level, the result is Pareto superior to the usual tacit agreement used by Benoit & Krishna (1987) and Davidson & Deneckere (1990), because firms remain cost efficient;

(2) An increase in the number of firms can lead to increased cost inefficiency;

(3) An increase in the number of firms can decrease welfare.

5. Conclusions

This paper supports the traditional view about excess capacity: excess capacity hinders collusion. The existing excess capacity results of Benoit & Krishna (1987), Davidson & Deneckere (1990) and Osborne & Pitchik (1987) are a direct consequence of the fact that firms distribute the rewards of colluding according to each firm's capacity. Therefore, firms engage in rent-seeking behaviour, installing larger capacities to 'steal'

market share from competitors. The inevitable result is over-investment in capacities, even if firms can tacitly collude in capacities, too (Benoit & Krishna 1987).

I avoid this rent-seeking behaviour by assuming that firms tacitly agree to produce the same output, regardless of how much capacity they installed. This is not explicit collusion, since the tacit contract has to be self-enforcing, and therefore subject to incentive constraints. Since increasing capacity does not increase a firm's share of production, the only strategic role for capacity here is to facilitate collusion.

I find that in such a context, firms would choose to restrict their capacity, as a commitment to collude. The less capacity they have, the less profitable it is to deviate and therefore firms can tacitly agree to produce a lower output. The incentive to decrease capacity is limited by the increased production inefficiency and also by the fact that if a firm decreases its capacity by too much, its competitor might prefer a non-cooperative outcome, when it has installed a much higher capacity.

I find that an increase in the number of firms (as long as firms are able partially collude) leads to aggregate capacity reductions and, therefore, is associated with declining production efficiency. It can happen that the decreased production efficiency overwhelms the increase in consumers' surplus, leading to welfare decrease. Indeed, even consumer surplus can decrease with an increase in the number of firms. This occurs when the number of firms increases from 6 to 7 in Figure 23.

An increase in the wage rate can switch the industry from a tacit collusion regime to a non-cooperative one, increasing welfare. This is a direct consequence of the main disadvantage (for the colluding firms) of the egalitarian agreement: it cannot be self-enforcing for all parameter ranges.

The use of egalitarian agreements by firms seeking collusion has a few antitrust implications, the most obvious of them being that, in some circumstances, insufficient capacity can be an indicator of collusion. Furthermore, an entry promoting policy may actually decrease welfare if it fails to substantially increase the number of firms in the market.

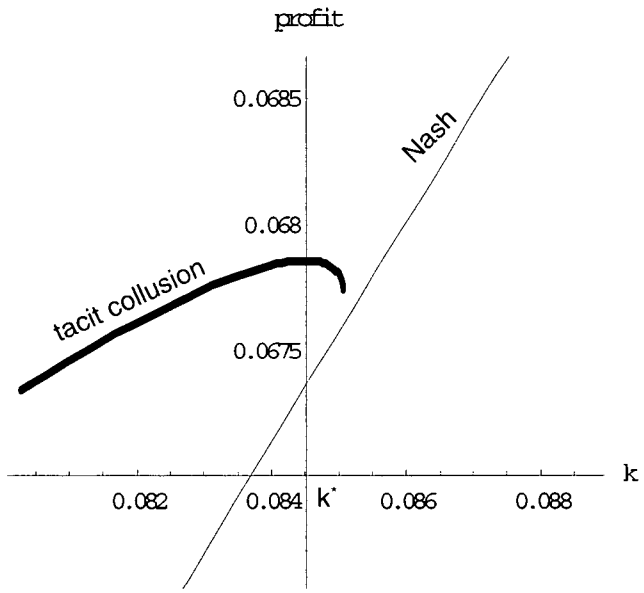

FIGURE 13. The tacit collusion equilibrium is only a local maximum ($\delta = 0.05, r = w = 0.1$)

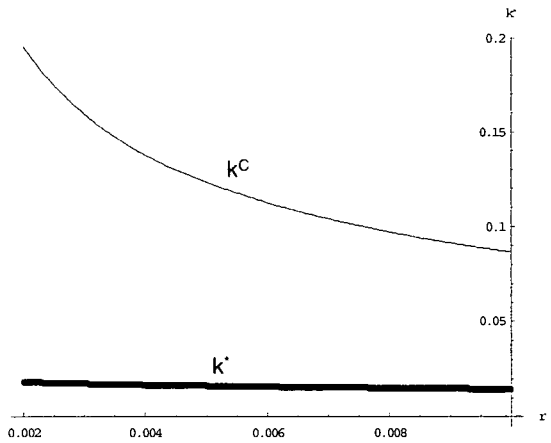

FIGURE 14. The effect of increases in the rental rate on equilibrium capacity, k^*, and cost-minimizing capacity k^C ($\delta = 0.31, w = 0.001$)

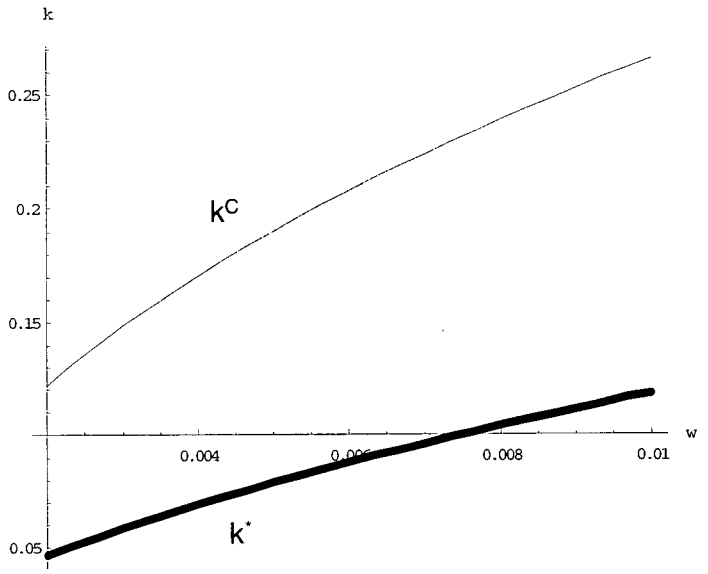

FIGURE 15. The effect of increases in the wage rate on equilibrium capacity, k^*, and cost-minimizing capacity k^C ($\delta = 0.35$,$r = 0.01$)

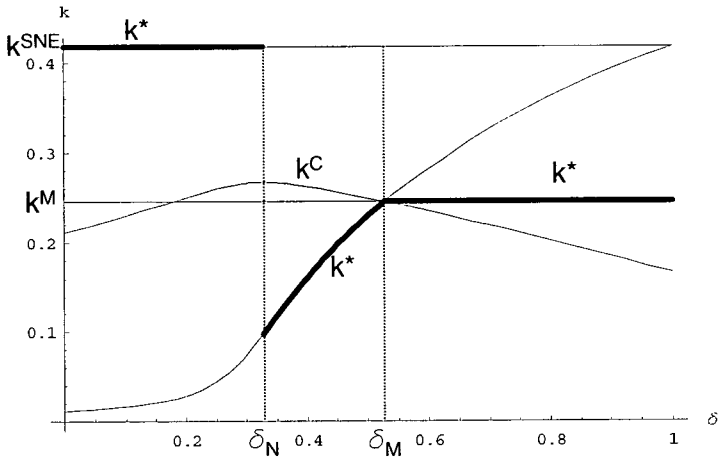

FIGURE 16. Comparative statics of k^* with respect to the discount factor ($r = w = 0.01$)

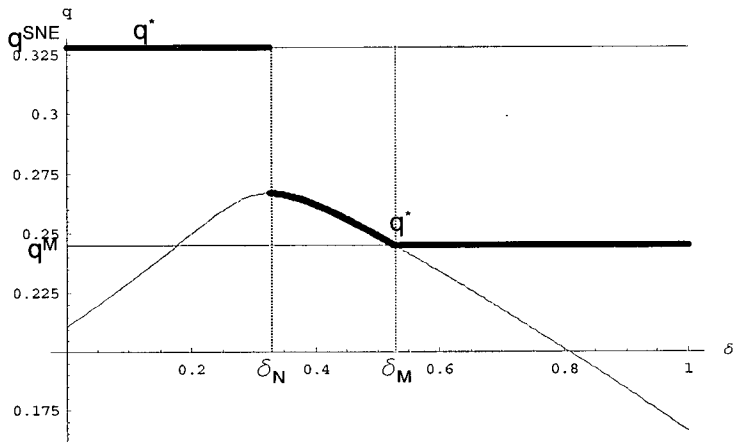

FIGURE 17. Profit maximizing output of a firm as a function of the discount factor ($r = w = 0.01$)

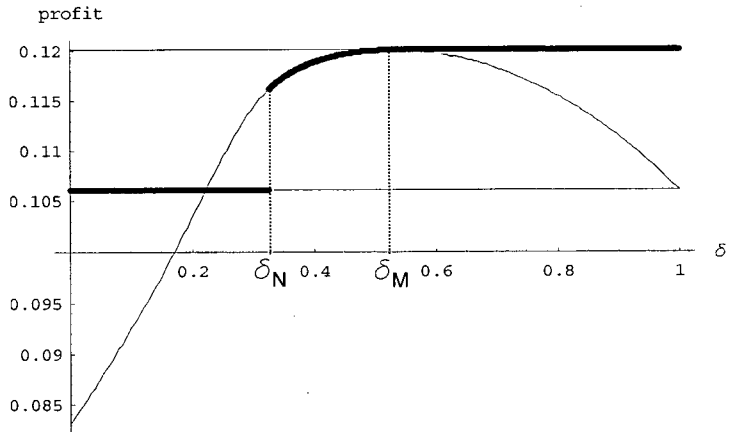

FIGURE 18. Profit of a duopolist as a function of the discount factor ($r = w = 0.01$)

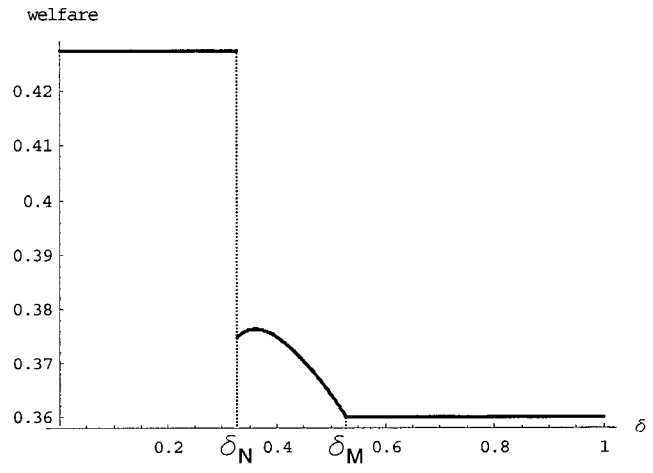

FIGURE 19. Welfare per period as a function of the discount factor ($r = w = 0.01$)

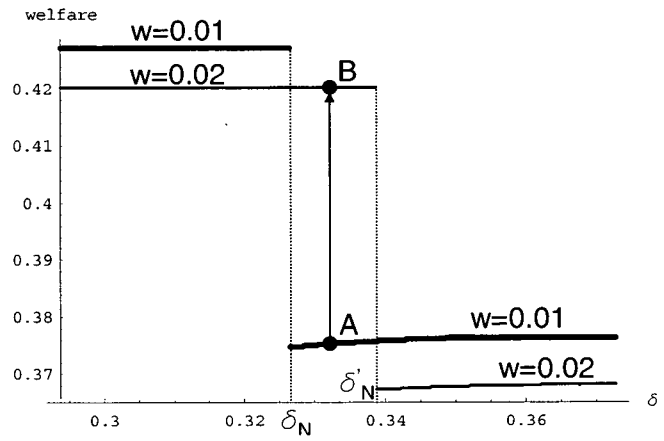

FIGURE 20. An increase in w increases welfare by changing δ_N to δ'_N, moving the industry from point A to B ($r = 0.01$)

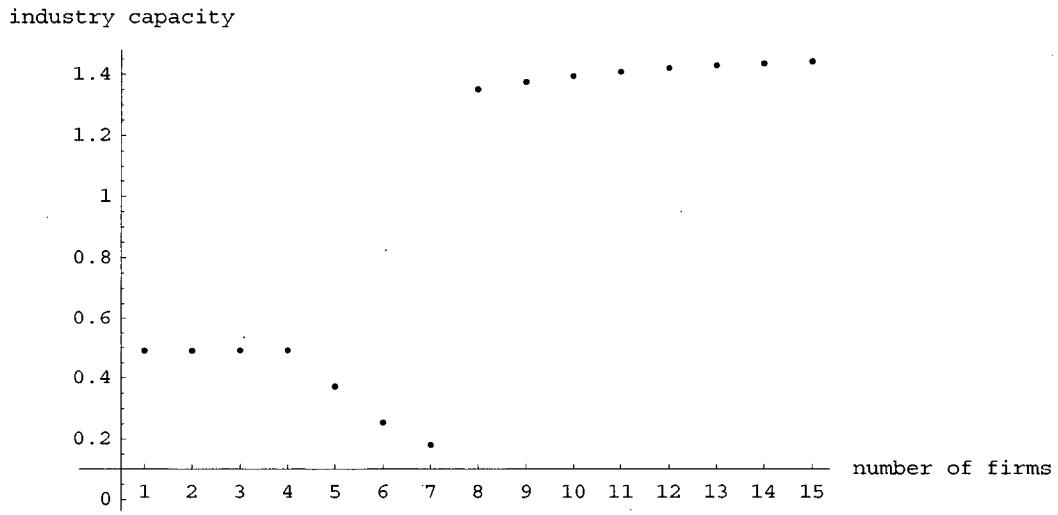

FIGURE 21. Industry capacity as a function of the number of firms. For $n = 1, 2, 3, 4$, firms replicate the monopoly outcome; tacitly collude partially for $n = 5, 6, 7$; and behave as Cournot firms for higher values of n. Industry capacity eventually declines (not shown) for even higher values of n. ($\delta = 0.6$, $r = w = 0.01$)

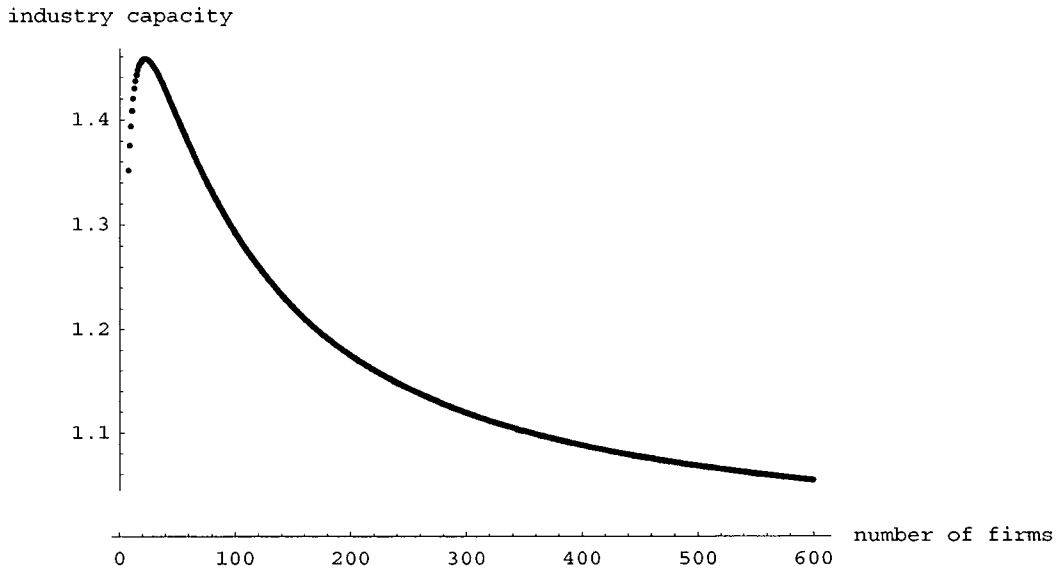

FIGURE 22. When firms cannot collude, industry capacity eventually declines for a higher number of firms. ($\delta = 0.6$, $r = w = 0.01$)

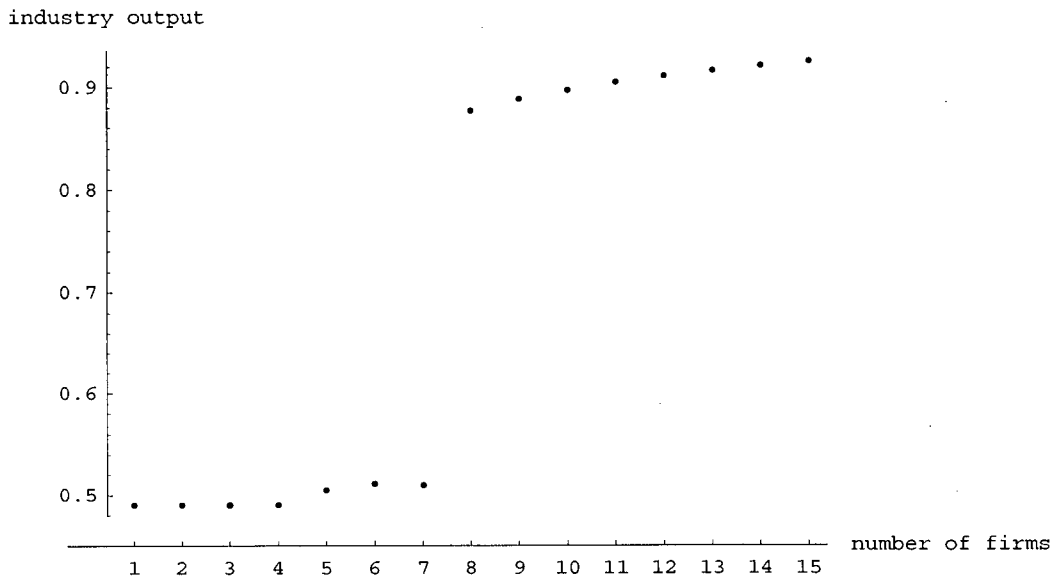

FIGURE 23. Industry output as a function of the number of firms. For $n = 1, 2, 3, 4$ firms replicate the monopoly outcome; tacitly collude partially for $n = 5, 6, 7$; and behave as Cournot firms for higher values of n. Note than the aggregate output for $n = 7$ is slightly lower than that for $n = 6$. ($\delta = 0.6$, $r = w = 0.01$)

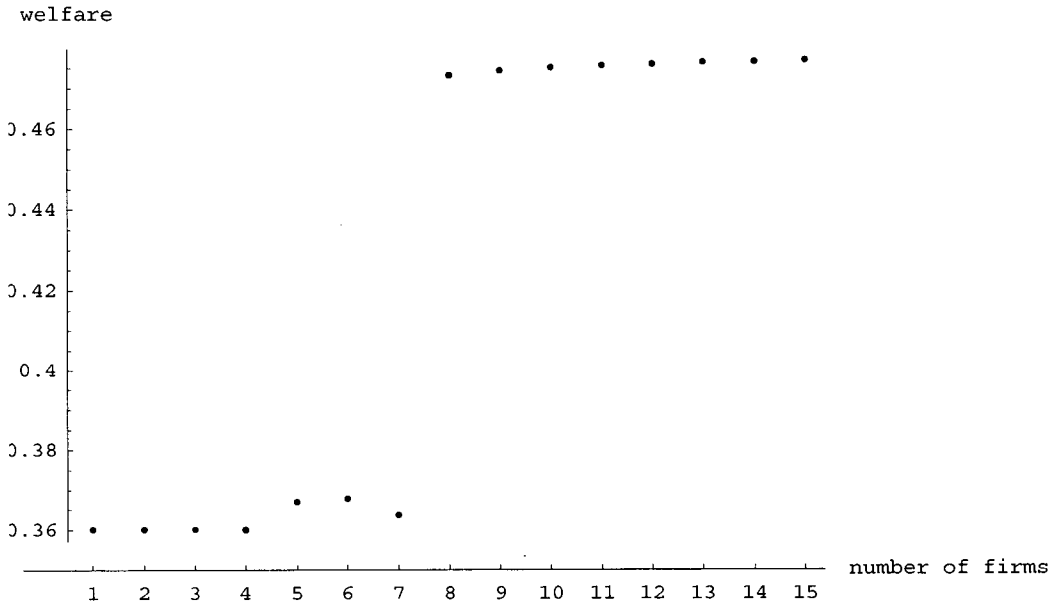

FIGURE 24. Welfare as a function of the number of firms. For $n = 1, 2, 3, 4$ firms replicate the monopoly outcome; tacitly collude partially for $n = 5, 6, 7$; and behave as Cournot firms for higher values of n.($\delta = 0.6$, $r = w = 0.01$)

CHAPTER 4

Competition in regional and national markets

This chapter examines the interaction between firms in two separate markets, connected indirectly by the fact that only one firm (called national firm) competes in both markets. Entry in one market affects the national firm's profitability of undertaking marginal cost reducing activities (such as R&D), and therefore affects the profitability of the firms in the other market. One surprising result of this model is that subsidizing entry in one market may motivate the national firm to deter entry in both markets, but does not necessarily reduce welfare.

1. Introduction

The recent deregulation of industries like telecommunications and air transportation raises, among other questions, the issue of who benefits from the deregulation in a country or market and how/whether these benefits spread beyond the originating market. This paper analyzes the effect of entry on the profitability of firms in separated markets that are indirectly connected by one large firm competing in both of them, while the other firms operate in only one of the markets.

This scenario could apply to industries like air transport (national carrier competing with local carriers), telecommunications (a firm like Telus competing as local exchange carrier in both BC and Alberta), or internet service providers (national providers competing with many local providers in many geographical areas). It has a special relevance to the Canadian airline industry, which became very concentrated in the aftermath of deregulation. According to the *The Economist*[1]:

> The Air Canada family, which includes various regional carriers as well as CAI[2], sells about 80% of domestic airline tickets and takes in about 90% of

[1] "Canada's not-so-open skies", in "*The Economist*", issue of September 2nd—8th, 2000, page 59.

[2] Canadian Airlines International, bought by Air Canada in December 1999.

> ticket revenues. Of the top 200 domestic routes, over half have no carrier other than Air Canada or one of its subsidiaries.

Air Canada competes locally with some small air transport companies, the most notable of them being a Calgary-based carrier, WestJet. According to the same newspaper, Air Canada is still the dominant domestic carrier, at least for the near future:

> WestJet, the charter carriers[3] and other small fry together account for less than a quarter of seat capacity. Even if they all follow through on expansion plans, their combined fleet will grow only from 75 to 137 aircraft by 2005, compared with Air Canada's projected fleet of 462.

The separated (interchangeably called *regional*) markets are linked indirectly by the dominant (alternatively called *national*) firm. Entry in one regional market affects decisions taken by the national firm (for example, its R&D effort) by changing the marginal benefit of its activities. In turn, these changes in the national firm's actions could affect competition in other regional markets in which entry did not take place, but possibly encouraging new entrants.

In an early paper, Bulow et al. (1985) considered another mechanism which can link demand-independent markets: increasing marginal costs. Shocks in the profitability of a market cause a firm to change its output in that market (direct effect), knowing that this affects its marginal cost (and, therefore, competition) in another market (indirect effect). More recently, in the context of rising marginal costs, Chen & Ross (2002) analyze the anticompetitive impact of multimarket contact by comparing a scenario with a national firm competing with two regional firm to one with two national firms in multimarket contact. I will show that one of their results, an apparent predatory pricing in one market associated with recoupment by charging higher prices in the other market, can be obtained also by the linking mechanism I am proposing in this paper.

The term 'regional' should not be interpreted only in its geographical meaning: the analysis presented here it is likely to be valid for two differentiated products that share a common stage of production. For example a 'regional' market could be the market for operating systems for desktop computers, overwhelmingly dominated by Microsoft with its range of Windows 9x/ME products, and the second regional market can be considered the market for operating systems for server computers, where Microsoft competes with its Windows 2000/NT against Sun (Solaris) and Linux, among others.

I show that if the number of firms in regional markets is exogenous, entry in one market might increase or decrease the intensity of national firm's efforts, depending on how expensive the

[3]Canada 3000, Air Transat and Royal. There are two newcomers contemplating the market, CanJet and RootsAir.

cost-reducing activity is. This result is consistent with those obtained by Delbono & Denicolo (1991) using a patent-race model. If research is relatively cheap, entry intensifies the cost-reducing activities; otherwise entry decreases the profitability of R&D and the marginal cost of the national firm increases. Ultimately, when the number of firms is large, entry decreases incentives to lower marginal cost.

If the number of firms is endogenous and the national firm is allowed to deter entry, it can do so by decreasing its marginal cost to the point where its rivals' profit becomes zero. The decision to deter entrants in a market is tightly connected with what happens in the other market. I will show that if the fixed cost in the market with the smaller fixed cost decreases, it can induce a switch from a regime in which entrants are deterred in both markets to a regime in which entrants are accommodated in both markets. A reduction in the fixed cost in the market with a higher fixed cost can produce deterrence in the other market. The welfare analysis yields ambiguous results: while deterrence decreases industry's profit, it could possibly improve welfare by increasing consumer's surplus. If R&D is sufficiently expensive, an entry-promoting policy can improve welfare. This ambiguous result is similar to that of Milgrom & Roberts (1982): entry deterrence is not necessarily socially harmful.

2. The Model

The model captures a two stage game. In the first stage the dominant firm chooses the level of R&D expenditure in order to decrease its marginal cost of production. This is an irreversible decision.

In the case of an air transportation firm, this R&D activity can be interpreted as anything that would reduce its marginal costs: purchasing more economical planes, better management of inventories, upgrading its fleet and increasing maintenance activities. Or it could decrease the quality of its services but compensate for it by increasing advertising to avoid a decrease in the number of customers. Increasing airport capacity and modernizing ground facilities cuts turnaround times and increases aircraft utilization, thereby decreasing the marginal cost.

In the example of computer operating systems, Microsoft can decrease its marginal cost (which for software consists mainly of the cost of providing technical support) by developing a product requiring less assistance in using it: that is, improving the quality of its software. As long as Windows 9x/ME share common parts with 2000/NT (an obvious example is the user interface and another the fact that many programs can be used on both platforms) improving

quality is going to be reflected in better operating systems (low technical support cost) for both desktop and server software. Another way to decrease the marginal cost is to invest in integrating the desktop and server software, and this is one of the targets of Microsoft, which it hopes to be accomplish with the Windows XP.

Alternatively, it can be imagined that lower marginal costs are due to economies of scale, but that cannot be exploited without investing in integrating different components of the firm (think of two firms merging and the cost required to integrate their business to be able to better reap the benefits of scale). If the national firm has many separate points of presence (plants or selling outlets), the economies of scale can be better exploited by better coordination among them. For example, a firm like 7/Eleven uses a proprietary software to manage the inventories of its outlets and deal with a reduced shelf space. This coordination or integration expenditures will be labeled R&D expenditures in this paper.

In the second stage, the national firm competes in Cournot fashion with firms in two separate markets. The national firm is the only one competing in both markets while the others compete in only one of the regional markets and they do not engage in R&D activities. This is a reasonable assumption since some investments are available only to large firms. In the previous example, large firms have larger scope for integration/coordination investments. Smaller firms replicate those activities, but with much lower benefit: lacking an appropriate size, economies of scale are likely to be small, even with large R&D investments. Even if they would undertake R&D, it is likely they will invest much less than the national firm, and I believe the qualitative conclusions of this paper will remain valid.

The aim of the model is to examine how a change in the number of firms in one market or the cost of operating in that market affects the firms in the other market in equilibrium.

2.1. Competition between n firms in a single market. This section solves for the second stage of the model. Firm 1 (i.e. the national firm) has already chosen its marginal cost c in the first stage. Since regional markets are separated and firms are assumed to have constant marginal cost, the competition in each market can be studied separately. Therefore in the rest of the section I will consider only one of the regional markets.

I use a linear inverse demand $P = a - bQ$, where Q is the total production of all the firms in the market, P is the price of output and a, b are strictly positive parameters (the two regions

are assumed to have identical demand curves). Firms are competing in Cournot fashion. Firm 1 has a constant marginal cost of c and all other firms have a constant marginal cost of c_0.

The profit (gross of R&D cost) of Firm 1 is given by:

$$\pi_1 = (P - c)q_1 = [a - b(q_1 + q_{-1}) - c]q_1 \tag{2.1}$$

where P is the price of the final output, q_1 is the output of Firm 1 and q_{-1} is the total output of all other firms in the market.

The profit of one of the (non-researching) regional firms is given by:

$$\pi_i = (P - c)q_i = [a - b(q_i + q_{-i}) - c_0]q_i \tag{2.2}$$

where q_i is the output of Firm i, q_{-i} is the total output of all other firms and $i \in \{2, \ldots, n\}$.

The first order conditions are the following:

$$\begin{aligned} bq_1 &= a - bQ - c \\ bq_2 &= a - bQ - c_0 \\ &\cdots \\ bq_n &= a - bQ - c_0 \end{aligned}$$

Adding all of the above, one immediately obtains the following expressions characterizing the Cournot equilibrium:

$$Q = \frac{na - c - (n-1)c_0}{b(n+1)} \tag{2.3}$$

$$P = \frac{a + c + (n-1)c_0}{n+1} \tag{2.4}$$

$$q_1 = \frac{a + (n-1)c_0 - nc}{b(n+1)} \tag{2.5}$$

$$q_i = \frac{a + c - 2c_0}{b(n+1)}, \quad i \in \{2, \ldots n\} \tag{2.6}$$

$$\pi_1 = bq_1^2 \tag{2.7}$$

$$\pi_i = bq_i^2, \quad i \in \{2, \ldots n\} \tag{2.8}$$

2.2. Competition in two markets. This section solves the first stage of the model, in which the national firm chooses its marginal cost by setting a level of R&D expenditures, taking into account how it affects the Cournot equilibrium of the second stage. In other words, I compute the Subgame Perfect Nash Equilibrium.

Assume that in market 1 there are n_1 firms and in market 2 there are n_2 firms. Firm 1 in market 1 is the same firm as the Firm 1 in market 2, and therefore maximizes the sum of its profits from both markets. As in the previous section, all other firms have constant marginal cost c_0. Firm 1 is the only one able to do R&D and is able to reduce its marginal cost to c at a cost of $\theta(c_0 - c)^2$, with $0 < c < c_0$ and where θ is a positive parameter capturing the difficulty of undertaking cost-reducing research. The relevant range of the R&D cost is $(0, \theta c_0^2)$.

The net profit of Firm 1 is given by:

$$\Pi_1 = \pi_{11} + \pi_{12} - \theta(c_0 - c)^2 = b(q_{11}^2 + q_{12}^2) - \theta(c_0 - c)^2 \tag{2.9}$$

where π_{1j} (q_{1j}) is the Cournot equilibrium profit (quantity) of firm 1 in market j.

As one can see, a change in b is equivalent to a change in θ and, therefore, I will normalize b to 1.

The first order condition for c is:

$$2(q_{11}\frac{\partial q_{11}}{\partial c} + q_{12}\frac{\partial q_{12}}{\partial c}) + 2\theta(c_0 - c) = 0 \tag{2.10}$$

or

$$-q_{11}\frac{n_1}{n_1 + 1} - q_{12}\frac{n_2}{n_2 + 1} + \theta(c_0 - c) = 0 \tag{2.11}$$

or

$$\frac{n_1}{(n_1 + 1)^2}[a + (n_1 - 1)c_0 - n_1 c] + \frac{n_2}{(n_2 + 1)^2}[a + (n_2 - 1)c_0 - n_2 c] = \theta(c_0 - c) \tag{2.12}$$

The second order sufficient condition is:

$$(\frac{n_1}{n_1 + 1})^2 + (\frac{n_2}{n_2 + 1})^2 - \theta < 0, \tag{2.13}$$

and, for any n_1 and n_2, is satisfied for sufficiently large θ. If the cost of research is too low (low θ), then Firm 1 would like to reduce its marginal cost to the minimum (i.e. zero), and there will be a corner solution $c^* = 0$. This is specific to this model, as a direct consequence of using a cost of research that is quadratic in c and a linear demand that yields profits that are also quadratic in c. Since this is an artificiality due to the functional forms used in the model (it is easier to imagine that the research cost for zero marginal cost should be rather enormous, if not infinitely large), I am only noting the possibility of corner solutions but I will consider only

interior solutions from now on. Then, the relevant range of θ is:

$$\theta \in \left(\left(\frac{n_1}{n_1+1}\right)^2 + \left(\frac{n_2}{n_2+1}\right)^2, +\infty \right) \tag{2.14}$$

Assuming an interior solution, equation (2.12) can be solved for c, yielding:

$$c^*(n_1, n_2, \theta) = \frac{\theta c_0 - \frac{[a+c_0(n_1-1)]n_1}{(n_1+1)^2} - \frac{[a+c_0(n_2-1)]n_2}{(n_2+1)^2}}{\theta - \frac{n_1^2}{(n_1+1)^2} - \frac{n_2^2}{(n_2+1)^2}} \tag{2.15}$$

The derivative of c^* with respect to n_2 is a very complicated expression, but it can be shown that:

$$sign[\frac{\partial c^*}{\partial n_2}] = sign[-(\frac{n_1}{n_1+1})^2 - (\frac{n_2}{n_2+1})^2 - \theta(1-n_2)] \tag{2.16}$$

If the second order sufficient condition (2.13) is satisfied then the first two terms of the right hand side of equation (2.16) are negative and the third is positive. Consequently, c^* can increase or decrease with n_2, depending on the value of θ. The value of θ at which $\frac{\partial c^*}{\partial n_2} = 0$ is:

$$\theta^c = \frac{n_2+1}{n_2-1}\left((\frac{n_1}{n_1+1})^2 + (\frac{n_2}{n_2+1})^2 \right) \tag{2.17}$$

Treating the number of firms as a continuous variable, I derive the following result:

PROPOSITION 2.1. *For an interior solution for Firm 1's R&D expenditure and for any market configuration (i.e. for any pair (n_1, n_2) with $n_2 > 1$), there is critical value of θ, denoted by θ^c, such that:*

(1) *the national firm's marginal cost of production decreases with the number of firms n_2 if $\theta < \theta^c$;*

(2) *the national firm's marginal cost of production increases with the number of firms n_2 if if $\theta > \theta^c$.*

(3) *If the national firm is a monopoly in market 2 (i.e. $n_2 = 1$) then $\frac{\partial c^*}{\partial n_2} < 0$ for all values of θ.*

To obtain the intuition for Proposition 2.1 one needs to recognize that the marginal cost plays a dual role in national firm's profit maximization problem: one is simply to minimize the cost of production and the other is to manipulate the national firm's reaction function. By choosing the marginal cost irreversibly in the first period, Firm 1 can commit to any marginal cost during the second period's Cournot competition. The more difficult the research is the more expensive it is to commit to a low ex-post marginal cost for strategic purposes: expensive

research reduces the strategic role played by the marginal cost. Since more entry means a lower output, reducing the marginal cost is less profitable for the national firm. In particular, if the national firm is a monopoly, its marginal cost does not play any strategic role. In this case entry is an event that triggers the transition from non-strategic to strategic behaviour. At the transition point it is always profitable to reduce costs, no matter how expensive research is.

As one can note, Proposition 2.1 ignores the integer constraint for n_1 and n_2. The full effect of entry cannot be evaluated by simply looking at the sign of the derivative $\frac{\partial c^*}{\partial n_2}$: even if that is negative, it does not say that the marginal cost is decreasing if the number of firms increases by a discrete amount of one. I now address this issue. Denote by Δ the change in Firm 1's equilibrium marginal cost when one more firm enters market 1:

$$\Delta \equiv c^*(n_1, n_2, \theta) - c^*(n_1, n_2 + 1, \theta) \tag{2.18}$$

It can be proved that:

$$sign[\Delta] = sign[n_1^2(1+4n_2+2n_2^2)+n_1(1+6n_2+2n_2^2)-(1+n_2)n_2-\theta(n_1^2+n_1-1)(1+n_2^2)^2] \tag{2.19}$$

PROPOSITION 2.2. *For an interior solution for Firm 1's R&D expenditure and for any market configuration, there is critical value of θ, denoted by θ_Δ^c, such that entry by another firm in market 2:*

(1) *intensifies Firm 1's R&D activities and decreases its marginal production cost for $\theta < \theta_\Delta^c$;*

(2) *lowers Firm 1's R&D expenditure and raises its marginal production cost for $\theta > \theta_\Delta^c$.*

Since Proposition 2.2 is the discrete version of Proposition 2.1 it is based on the same intuition. However it adds the feature that even in the case of the qualitative transition from monopoly to duopoly, the discrete change may be enough for expensive research to counterbalance the marginal cost's strategic role. Proposition 2.2 is not trivial, as a casual look at equation (2.19) would suggest: the last term in the right hand side of the equality (the only term that can make negative the equality) cannot be infinitesimally small since the second order solution (2.13) imposes a lower limit on the value of θ.

For θ slightly higher that θ_Δ^c, the marginal cost of production of the national firm increases by little as response to entry in market 2. This increase in the marginal cost might be insufficient to counterbalance the increased number of firms in market 2 and therefore the equilibrium price

will fall in market 2. The increase in marginal cost is not accompanied by a change in the number of firms in market 1 and therefore the equilibrium price will increase in this market. This behaviour might look as if the national firm charges a predatory price in market 2 while trying to recoup its losses by increasing the price it charges in market 1, a result similar to one obtained by Chen & Ross (2002).

It would be useful to separate the strategic effect of lowering rivals' output from the non-strategic effect (pure cost minimization) on the choice of the marginal cost. The easiest way to do this is to remove the commitment ability of the national firm. This is done here by eliminating the first stage and letting the firm make simultaneous decisions on the quantities to sell in regional markets and the level of marginal cost. In other words, the national firm would solve the following problem, if it were to behave "non-strategically"[4]:

$$\max_{\{q_{11},q_{12},c\}} \left[a - c - b(q_{11} + q_{-11})\right] q_{11} + \left[a - c - b(q_{12} + q_{-12})\right] q_{12} - \theta(c_0 - c)^2 \tag{2.20}$$

Regional firm i in market j's problem remains the same as before:

$$\max_{\{q_{ij}\}} \left[a - c_0 - b(q_{ij} + q_{-ij})\right] q_{ij} \tag{2.21}$$

where $j = 1, 2$ and $i = 2, \ldots, n_j$.

The solution of the above problem defines the optimal non-strategic marginal cost, denoted by $\overline{c}$:

$$\overline{c}(n_1, n_2, \theta) = \frac{2\theta c_0 - \frac{a+c_0(n_1-1)}{(n_1+1)} - \frac{a+c_0(n_2-1)}{(n_2+1)}}{2\theta - \frac{n_1}{(n_1+1)} - \frac{n_2}{(n_2+1)}} \tag{2.22}$$

The optimal non-strategic marginal cost is increasing with the number of firms in any regional market. This is because the potential savings from reducing the marginal cost increase with the output and the national firm finds more profitable to economize on research when its market share declines. We may define the "strategic effect" as the difference between the optimal non-strategic marginal cost and the optimal marginal cost in the commitment case:

$$c^S \equiv \overline{c}(n_1, n_2, \theta) - c^*(n_1, n_2, \theta). \tag{2.23}$$

This strategic effect explains the non-monotonicity of Δ and $\frac{\partial c^*}{\partial n_2}$. The non-strategic marginal cost of Firm 1 increases with entry, as observed above. The strategic change in the marginal cost is negative with new entry if the number of firms is relatively small and research not too

[4]This might be an abuse of language, since firms are still competing strategically in quantities.

expensive, but positive if the number of firms is large. Thus, if research is sufficiently expensive, the marginal cost on balance is monotonically increasing in the number of firms in the market. When research is inexpensive, on the other hand, entry can affect Firm 1's marginal cost non-monotonically.

Note that the fact that the national firm has a lower marginal cost if it invests in R&D is not essential for the intuition behind these results, and it is used only to simplify the model. What it is essential is the ability of the national firm to decrease its marginal cost. The model can be applied even if $c > c_0$. An example for this situation would be the case of brand-name pharmaceutical firms (which have high marginal cost due to expenses to maintain the brand recognition) competing with low cost generic pharmaceutical firms.

3. Entry

In the previous section, the number of regional carriers was exogenously specified. In this section I endogenize the entry decision of regional firms by allowing the national firm to behave strategically.

To have limited entry, I have to assume there is a fixed cost of operating in the market. The model is still a two-stage one, but the number of regional firms is now endogenous. The national firm has already sunk the fixed cost and chooses irreversibly its marginal cost during the first stage, taking into account not only how it affects directly the second stage's Cournot competition but also how it affects the number of firms in the market. In the second stage potential entrants decide whether to enter the market, knowing that if they enter they have to incur a fixed cost and that they will compete in Cournot fashion with the incumbent firm.

To keep the model as simple as possible, I will assume that initially the national firm is a monopoly in both regional markets, and that in each regional market there is only one potential entrant that decides to enter if the profit is strictly positive. I will also allow the fixed cost in market i, denoted by F_i, to be different from that in the other regional market.

Based on the previous results from equations (2.3) to (2.8), I will define the following:

- $\pi_M(c) = \frac{(a-c)^2}{4b}$ the profit of an unconstrained monopoly in a regional market;
- $\pi_{DN}(c) = \frac{(a-2c+c_0)^2}{9b}$ the national firm's profit in a regional market if it is in a duopoly with marginal cost c and the rival has a marginal cost of c_0;
- $\pi_{DE}(c) = \frac{(a+c-2c_0)^2}{9b}$ the profit of an accommodated entrant with marginal cost c_0 when the national firm has marginal cost c.

If Firm 1 accommodates entrants in both markets it realizes a profit of

$$\Pi_{AA}(c) = 2\pi_{DN}(c) - \theta(c_0 - c)^2 \tag{3.1}$$

and the optimal value of c is $c^* = \frac{9bc_0\theta - 4(a+c_0)}{9b\theta - 8}$, yielding a total profit of

$$\widehat{\Pi}_{AA} = \frac{2(a-c_0)^2\theta}{9b\theta - 8} \tag{3.2}$$

If the national firm wants to deter entry in a market with F fixed cost for either potential entrant, it has to choose a marginal cost that is at least as low as the one that solves $\pi_{DE}(c_D) = F$:

$$c_D(F) = 2c_0 - a + 3\sqrt{bF} \tag{3.3}$$

Therefore if Firm 1 decides to deter in market i and to accommodate in market j, it makes a profit of

$$\widehat{\Pi}_{DA}(F_i) = \pi_{DN}(c_D(F_i)) + \pi_M(c_D(F_i)) - \theta[c_0 - c_D(F_i)]^2 \tag{3.4}$$

where it is assumed that $F_i > F_j$.

I will consider only the case $F_1 \geq F_2$. (The other case can be solved by simply switching markets in this exercise.) Note that if the national firm deters entry in market 2, it also automatically deters entry in market 1. Deterrence in both markets is obtained by choosing a marginal cost $c_D(F_2)$, and this yields the national firm a profit of

$$\widehat{\Pi}_{DD}(F_2) = 2\pi_M(c_D(F_2)) - \theta[c_0 - c_D(F_2)]^2 \tag{3.5}$$

Firm 1 has three options:

i) accommodate entry in both markets (denoted by AA);

ii) deter entry in both markets (denoted by DD);

iii) deter entry in market 1, accommodate entry in market 2 (denoted by DA).

To find the range of fixed costs for which Firm 1 accommodates both potential entrants, one has to determine the combinations (F_1,F_2) which satisfy the two inequalities:

$$\widehat{\Pi}_{AA} \geq \widehat{\Pi}_{DA}(F_1) \tag{3.6}$$

$$\widehat{\Pi}_{AA} \geq \widehat{\Pi}_{DD}(F_2) \tag{3.7}$$

Let $\widetilde{F}^c$ be the critical value of F_1 that makes firm 1 indifferent between accommodating or deterring in market 1, provided that it does not deter in market 2 (i.e. $\widetilde{F}^c$ solves $\widehat{\Pi}_{AA} = \widehat{\Pi}_{DA}(\widetilde{F}^c)$

). Let $\widehat{F}^c$ be the critical value of fixed cost of market 2 that makes the national firm indifferent between total accommodation and total deterrence (solves $\widehat{\Pi}_{AA} = \widehat{\Pi}_{DD}(\widehat{F}^c)$). In the space of fixed costs, the national firm accommodates both entrants in region AA defined by:

$$AA = \{(F_1, F_2) | F_1 \geq F_2; F_2 \geq 0; F_2 \leq \widehat{F}^c; F_1 \leq \widetilde{F}^c\} \tag{3.8}$$

The national firm deters entry in both markets if

$$\widehat{\Pi}_{DD}(F_2) \geq \widehat{\Pi}_{AA} \tag{3.9}$$

$$\widehat{\Pi}_{DD}(F_2) \geq \widehat{\Pi}_{DA}(F_1) \tag{3.10}$$

Equation (3.9) defines the same critical value $\widehat{F}^c$ as equation (3.7).

Equation (3.10) with equality defines an implicit relationship between F_2 and F_1:

$$F_2 = \phi(F_1) \tag{3.11}$$

Along the curve defined by equation (3.11) the national firm is indifferent between accommodating and deterring in market 2, provided that it deters in market 1. The profit of deterring in both markets depends only on F_2 since deterrence in market 2 implies deterrence in market 1. As long as entry in not blockaded (that is, when the national firm as a monopolist would choose a marginal cost that automatically deters entry), $\widehat{\Pi}_{DD}(F_2)$ is an increasing function of its argument. F_2 does not influence the profit of the national firm if entry is accommodated in market 2, and therefore $\widehat{\Pi}_{DA}$ does not depend on F_2. The only way to increase $\widehat{\Pi}_{DA}$ is to increase F_1. Therefore the function $\phi(F_1)$ is strictly increasing in its argument as long as entry is not blockaded in market 1 and becomes constant afterwards. Denoting by F_B the minimum value of fixed cost for which entry in market 1 is blockaded, the function $\phi(F_1)$ becomes constant for $F_1 \geq F_B$.

Above this curve the national firm prefers to deter in both markets (higher F_2 makes deterrence cheaper), and below it prefers to accommodate entry in market 2. Formally, the national firm deters entry in both markets in region DD of (F_1, F_2) space, defined by:

$$DD = \{(F_1, F_2) | F_1 \geq F_2; F_2 \geq \max(\widehat{F}^c, \phi(F_1))\} \tag{3.12}$$

In the remaining area of (F_1, F_2) space, denoted DA, the national firm deters entry in market 1 and accommodates entry in market 2:

$$DA = \{(F_1, F_2) | F_1 \geq F_2; F_2 \leq \widehat{F}_2(F_1); F_1 \geq \widetilde{F}^c\} \tag{3.13}$$

This partitioning of the fixed cost space is illustrated in Figure 25, where I include the symmetric case when $F_2 \geq F_1$.

Figure 26 shows the same partition for the case in which the national firm does not behave strategically with respect to R&D. As one can see by allowing the national firm to behave strategically, reduces the region in which both regional firms survive (dotted area in picture): a part is taken over by the region DD, and another by the regions in which only one regional firm survives (AD or DA).

PROPOSITION 3.1. *If $F_i \in (\widehat{F}^c, \widetilde{F}^c)$, with $i = 1, 2$, then a change of appropriate magnitude in the value of fixed cost in* only one *of the markets is enough to induce the national firm to abandon a policy of deterrence in both markets in the favor of a policy of accommodating entrants in both regional markets.*

If $F_2 \in \left(\widehat{F}^c, \widehat{F}_2(F_1)\right)$, then a decrease of sufficient magnitude in the value of fixed cost in market 1 will induce the national firm to stop accommodating the potential entrant in market 2. Under similar conditions, a decrease in the fixed cost in market 2 can induce deterrence in market 1.

The first part of Proposition 3.1 is illustrated in Figure 25 by a movement from point A to point B (exogenous decrease in cost of entry in market 2) and the second part by a movement from point C to point E (exogenous decrease in cost of entry in market 1). If F_1 is high, market 1 is very attractive for the national firm because it can charge a high price with little concern that this high price would induce entry and therefore there is not much use for the strategic component of cost reducing research. Therefore, the national firm could extract a high profit with a relatively high marginal cost, thus accommodating the entrant in market 2. If F_1 decreases, market 1 cannot be exploited as easily as before and it could be worthwhile to increase the cost-reducing research activities that can ultimately lead to deterrence of both regional firms.

If the government wants to stimulate entry by pursuing a policy of lowering fixed costs by subsidization, it might happen that lowering the fixed cost in only one market would induce the national firm to accommodate both entrants, switching from total deterrence to complete accommodation. Moreover, a policy of reducing the fixed costs to the same amount is likely to be the most expensive way of increasing competition in both markets: moving along the 45 degree line is a more expensive way to reach region AA. Worse even, trying to reduce the fixed cost in the market with higher fixed cost (that is, going from C to E in Figure 25) can produce the

opposite result, in which the national firm becomes a (constrained) monopoly in both regional markets.

The effect of entry in one market on the profit of regional firms in the other market is ambiguous. However, for reasons I have already discussed, the larger the number of firms already in place, the more likely is it that entry increases the national firm's marginal cost of production. This makes the regional firms in the other market better off. The strategic advantage of influencing the Cournot equilibrium becomes insignificant with a large number of entrants and the only motivation for decreasing the marginal cost is the usual one of cost-minimization. When the market share of Firm 1 decreases, it is not as profitable to reduce the marginal cost.

If the number of established firms is small, the research technology plays a crucial role: the cheaper the technology is, the more likely is it that entry in one market induces the national firm to become more competitive by strategically decreasing its marginal cost, and this makes the regional firms in the other market worse off.

4. Welfare Analysis

It is not obvious how the welfare generated by the markets changes when the economy moves, as a result of exogenous changes in fixed costs, from one region of the (F_1, F_2) space to another. (I still keep the assumption $F_2 < F_1$ unless otherwise specified.) An interesting question is whether a move from region DD to region AA constitutes a welfare improvement. Moving from DD to AA tends to increase competition and decreases prices, but society has now to incur the fixed costs for the two entrants. Moreover, increased competition has an ambiguous effect on the national firm's marginal cost. The Cournot equilibrium price, therefore, might increase with competition. As a result even consumers may be worse off with entry.

Deterrence decreases the industry's profit, in the sense that if the national firm is indifferent between deterring a potential entrant in a regional market then the industry profit with deterred entry is lower than industry profit with accommodated entry. This situation is represented by a dot on the borders between AA, DD, AD or DA in Figure 25. If the entrant is accommodated and since it is not blockaded (as defined by Dixit (1980)), it earns more than the fixed cost of entry and contributes positively to industry profit. If there is a competitor in the other market, it is going to be hurt by the decrease of the national firm's marginal cost required to drive the potential entrant out, and therefore the firm in the other regional market has lower profits due to entry deterrence in one market. The overall effect is that deterrence cannot increase welfare

by increasing industry's profit. If deterrence increases welfare, it has to do so by increasing consumer surplus by enough to compensate for the decrease in industry profit.

I will take the welfare function to be the sum of producer and consumer surplus[5]. I will denote the consumer surplus in market i by CS_i and it can easily be shown, in the case of linear demand, to be given by following formula:

$$CS_i = Q_i^2 = (a - P_i)^2, \tag{4.1}$$

where P_i, Q_i are the price and total output in market i.

Denoting by W_r, $r \in \{AA, DD, DA\}$ the welfare function in outcome r, we have:

$$W_{AA} = \widehat{\Pi}_{AA} + 2\pi_{DE}(c^*) - F_1 - F_2 + 2(\frac{2a - c^* - c_0}{3})^2 \tag{4.2}$$

$$W_{DD} = \widehat{\Pi}_{DD}(F_2) + 2(\frac{a - c_D(F_2)}{2})^2 \tag{4.3}$$

$$W_{DA} = \widehat{\Pi}_{DA}(F_1) + \pi_{DE}(c_D(F_1)) - F_2 + \\ +(\frac{a - c_D(F_1)}{2})^2 + (\frac{2a - c_D(F_1) - c_0}{3})^2 \tag{4.4}$$

The iso-welfare curves are lines with slope -1 in region AA because the Cournot equilibrium does not depend on either fixed cost. In region DD welfare depends solely on the smaller fixed cost in the two regional markets and therefore the iso-welfare curves are horizontal lines if $F_2 < F_1$ and vertical lines otherwise. In region DA, the iso-welfare lines are upward or downward-sloping lines, depending on θ.

To decide whether welfare improves when we move across regions in Figure 25, we have to examine closely the welfare along the borders (where the national firm is indifferent between deterring or accommodating). Along the borders a switch $AA \rightarrow DA$ or $DA \rightarrow DD$ is accompanied by a decrease in industry profit, as argued before. Such a switch is due to the national firm decreasing its marginal cost and this tends to increase consumer surplus. On the other hand, in the market where the entrant is deterred there is a monopoly and this tends to increase the price. The bigger θ is, the smaller the decrease in marginal cost accompanying the switch, and therefore more likely is it that the consumer surplus has decreased in the market where the potential entrant is deterred while increasing very little in the other regional market. This

[5]This standard welfare measure is reasonable when income effects are negligible and externalities are absent

intuition is supported by simulations shown in Figures 27 through 30. Figures 27 and 28 show that $W_{DD} > W_{AA}$ for low values of θ and $W_{DD} < W_{AA}$ for high values of θ at both ends of the boundary between AA and DD . Similarly, Figure 29 shows that a switch $DA \rightarrow DD$ is welfare improving for small θ but welfare reducing for large θ.

The analysis is slightly different for a switch $AA \rightarrow DA$ (Figure 30) . Along the vertical line $F_1 = \widetilde{F}^c$ in Figure 25, the difference $W_{AA} - W_{DA}$ is constant because it does not depend on F_2 and F_1 is fixed. If one of these two regimes is socially superior at one point along the border between AA and DA, then that regime is superior along the whole border. This kind of property does not hold along the border between regions AA and DD, as one can easily see by comparing Figure 27 and 28[6].

Figure 30 seems to suggest that along the border between regions AA and DA, the welfare is higher if the national firm accommodates entrants in both regional markets. The explanation might be the fact that the cost of entry in market 1 in this case ($\widetilde{F}^c$) is so high that the decrease in marginal cost accompanying deterrence does not sufficiently increase consumer surplus to increase the overall welfare. Note that a switch from DA to DD (movement from C to E in Figure 25) happens when the cost on entry in market 2 (the relevant market for the decision whether to deter or accommodate) is lower than $\widetilde{F}^c$, and therefore deterrence is accompanied by a larger drop in national firm's marginal cost, with the possibility of welfare increasing deterrence (at low values of θ in Figure 29).

In conclusion, we may say the following. If the R&D is sufficiently expensive (θ is high), deterrence is likely accompanied by a decrease in welfare. If outcomes are in the close vicinity of the border between the regions, policies moving the outcome from point A to point B in Figure 25 (by subsidizing the entry in market 2) or from point E to point C (by *taxing* or regulating entry in market 1) are welfare improving policies. The opposite happens if θ is sufficiently low. It is likely that policies inducing a switch from the DA to the AA region by subsidizing entry in market 1 improves welfare for all values of θ.

[6]For example, when $\theta = 3.6$, AA has higher welfare than DD around point $(\widehat{F}^c, \widehat{F}^c)$, in Figure 25, but the situation is reversed at the other end of the border, around point $(\widetilde{F}^c, \widehat{F}^c)$.

As mentioned before, these conclusions do not differ qualitatively from those obtained in a single market where there exists the possibility of deterrence: deterrence has ambiguous effect on welfare. Nevertheless, this analysis is necessary to draw conclusions about the welfare implications of policies switching between the different regimes in Figure 25.

5. Conclusions

For a firm that is able to do R&D (called here the national firm) to decrease its marginal cost of production, the marginal cost plays a dual role. The first is a non-strategic role, and consists in minimizing the cost of production for a given output. The second role is a strategic one because it allows the firm to choose its own reaction function.

With more firms in the market, the output of the national firm is lower and, therefore, reducing the marginal cost becomes less profitable, as long as the R&D cost is unrelated to the level of output. As a result, the non-strategic role diminishes with the number of firms.

When the number of firms is small, the actions of the R&D-enabled firm have a sizeable impact on other firm's actions. This makes possible, under certain conditions, for the strategic role to increase with the number of firms. Of course, as the number of firms increases, the impact of one firm's action spreads over more and more firms, and the impact of the national firm's action becomes negligible. Ultimately, the strategic role becomes insignificant too.

The endogenous marginal cost of Firm 1 provides an indirect connection between two separate (regional) markets since the national firm competes in both of them. Depending on the markets' configuration and research technology, entry in one of them affects the profitability of the regional firms in the other market. There is a fundamental tendency for entry to reduce the profitability of cost reducing activities and, therefore, to increase the marginal cost of production. But it can also happen that the strategic effect increases drastically with entry. With sufficiently cheap R&D technology, entry in one market makes firms in the other market worse off when there is a small number of firms in the entry market is small, but better off in the large number of firms case. In the small number case the market price in the market falls more than in the other market: the price falls in both markets because the national firms is a tougher competitor (smaller marginal cost) but it falls more in the entry market because of increased competition (larger number of firms). Even when entry softens the competition from the national firm, entry in one market might decrease the equilibrium price in that market while increasing the price in the other market. As a result, the national firm's behaviour might seem predatory to new

entrants (potentially 'recouping' losses by charging a higher price in the non-entry market). In the example of Canadian air transport industry new entrants WestJet and CanJet accused Air Canada of predatory pricing immediately after entering the market (McArthur 2000).

As is well known, a firm's ability to shift its own reaction function is an instrument for entry deterrence (Dixit 1980, Tirole 1988). This is just a way to commit the national firm to a limit pricing strategy. The fact that the national firm has to take into account both markets when deciding whether it should deter entry provides another link between regional markets. It may be that decreasing the fixed costs in only one market (through a provincial subsidy, say) is the cheapest way to achieve a higher number of firms in both markets. It can happen that decreasing the fixed cost in one market persuades the national firm to accommodate the competitor in that market, triggering accommodation in the other market, too. Subsidizing the market with the larger fixed cost, however, would yield no result other than a movement along a social indifference curve. A more expensive (for the government, or even for the society if taxes are distortionary) way to achieve the same results is to subsidize the fixed cost to the same level in both regional markets.

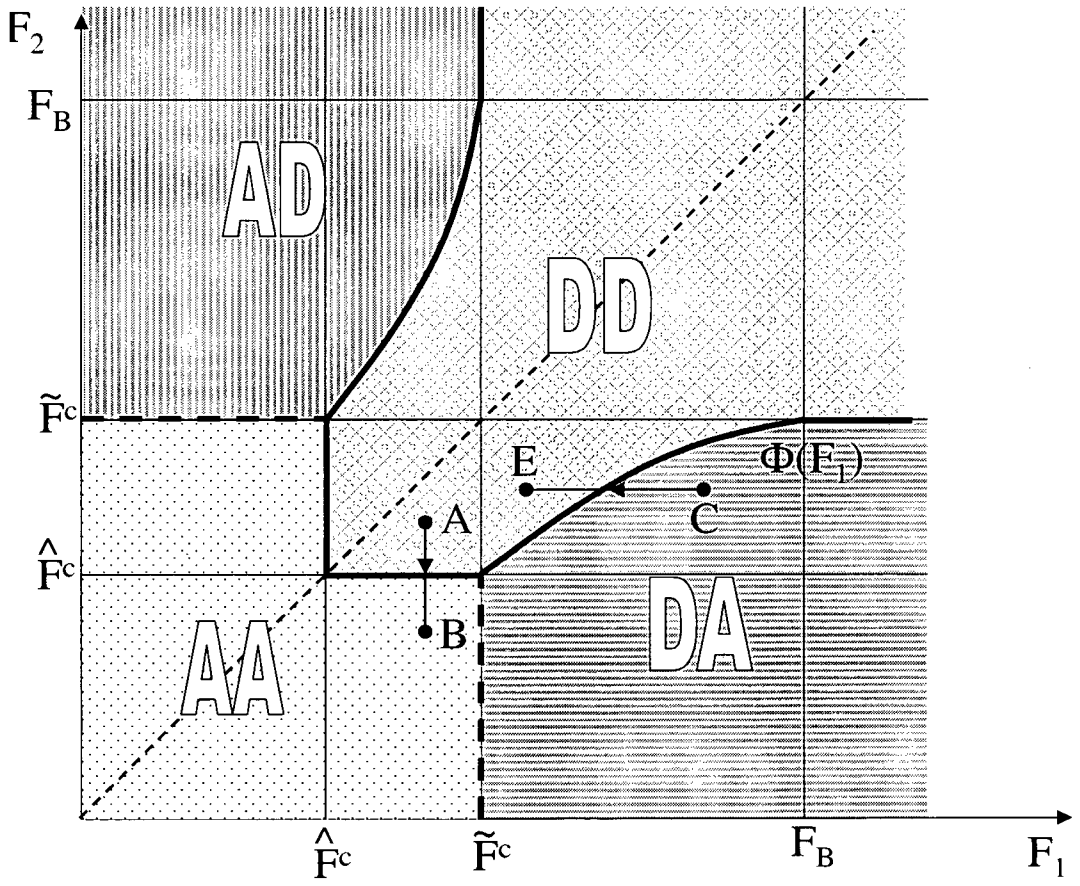

FIGURE 25. Cost of entry space partitioned according to national firm's decision to accommodate or deter entry.

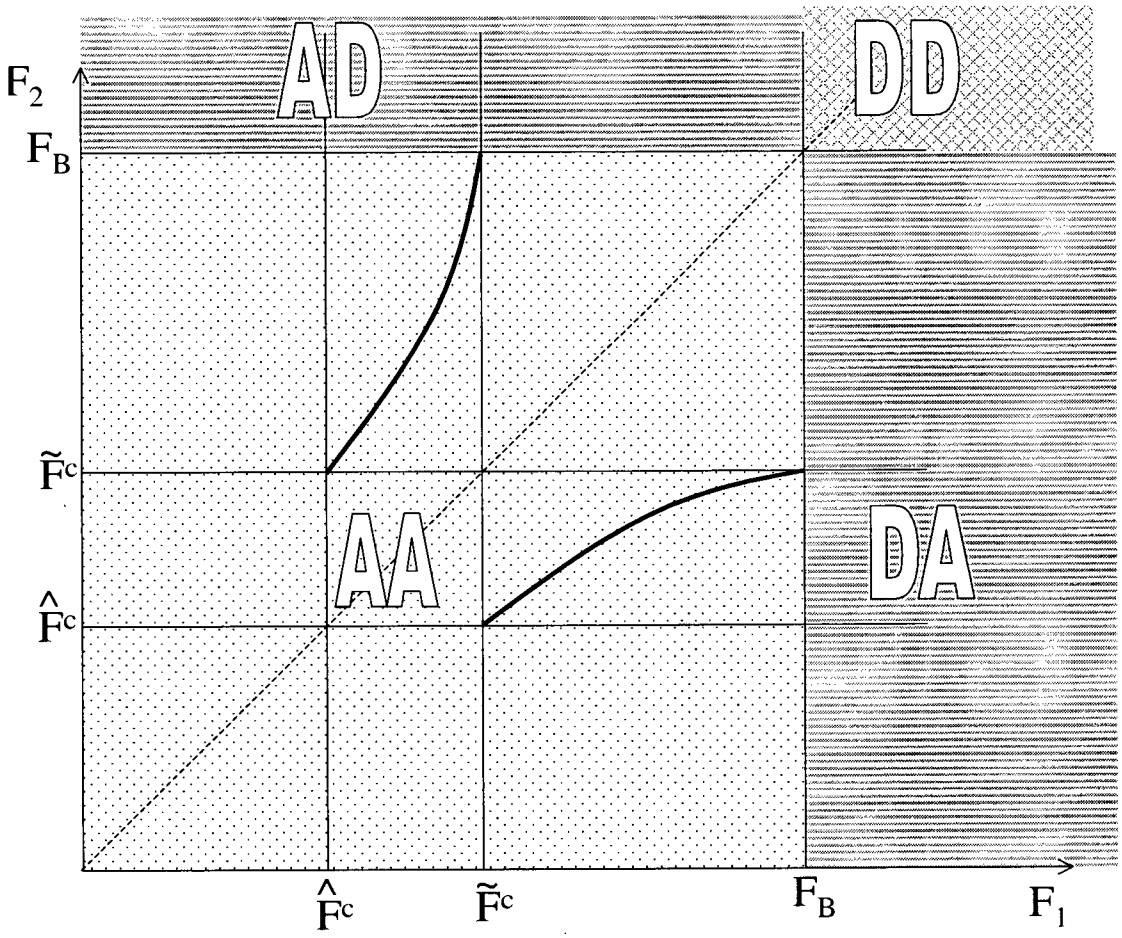

FIGURE 26. Cost of entry space partitioned according to whether entry is blockaded. The national firm does not behave strategically with respect to R&D. Compare with Figure 25 to see the difference between strategic and non-strategic behaviour.

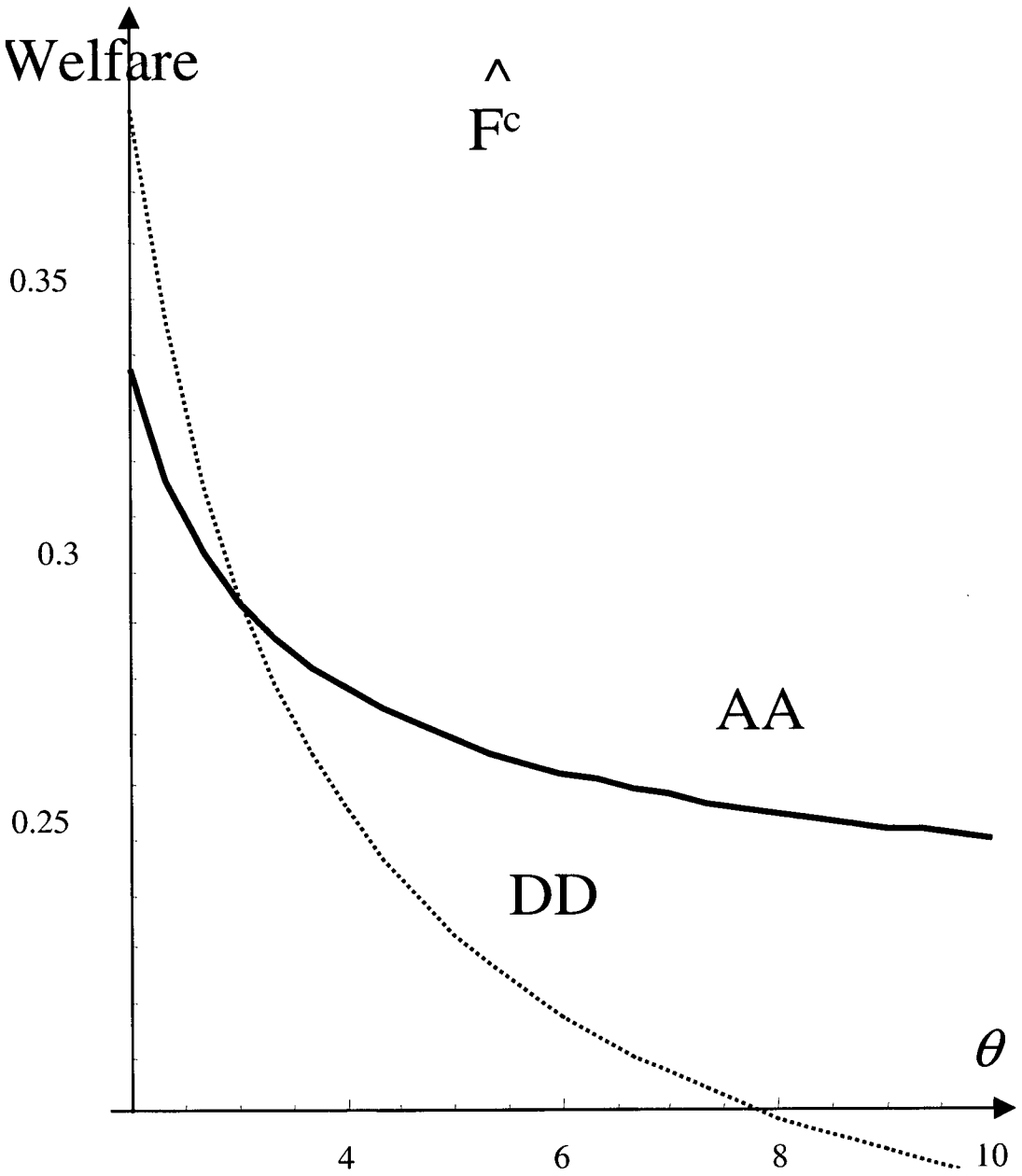

FIGURE 27. Welfare comparison between regimes AA and DD around the point $(\widehat{F}^c, \widehat{F}^c)$

Parameters: $a = 1$, $c_0 = 0.5$

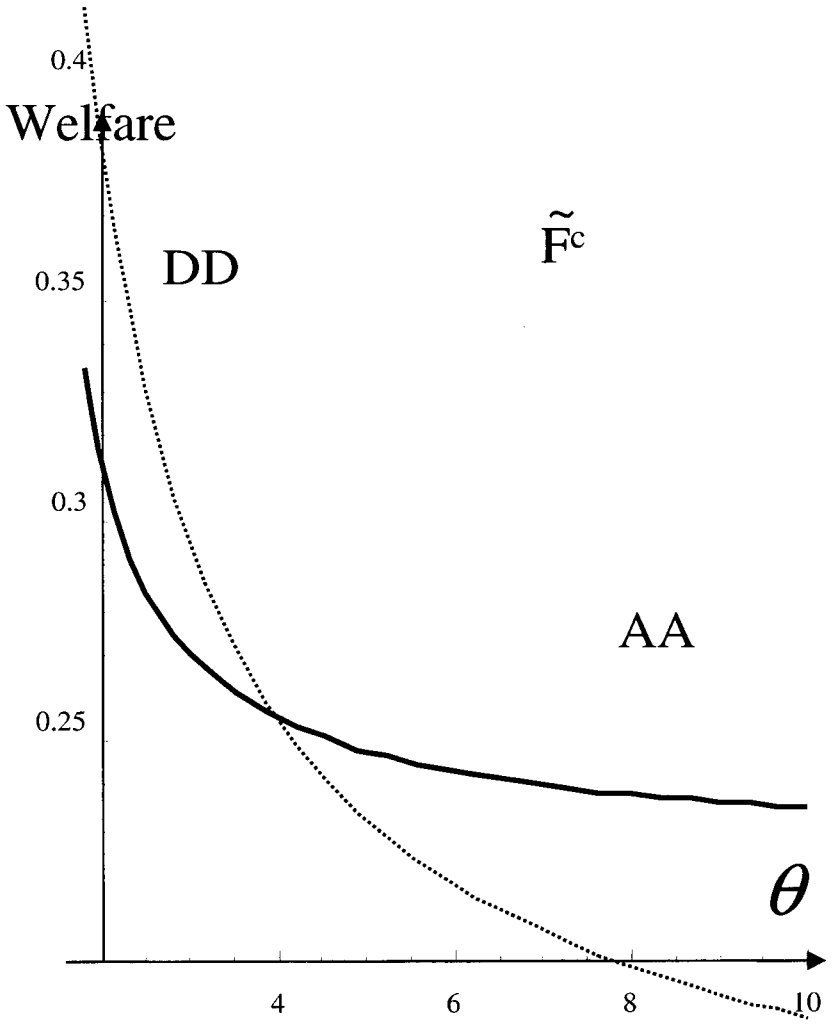

FIGURE 28. Welfare comparison between regimes AA and DD around the point $(\widetilde{F}^c, \widehat{F}^c)$

Parameters: $a = 1$, $c_0 = 0.5$

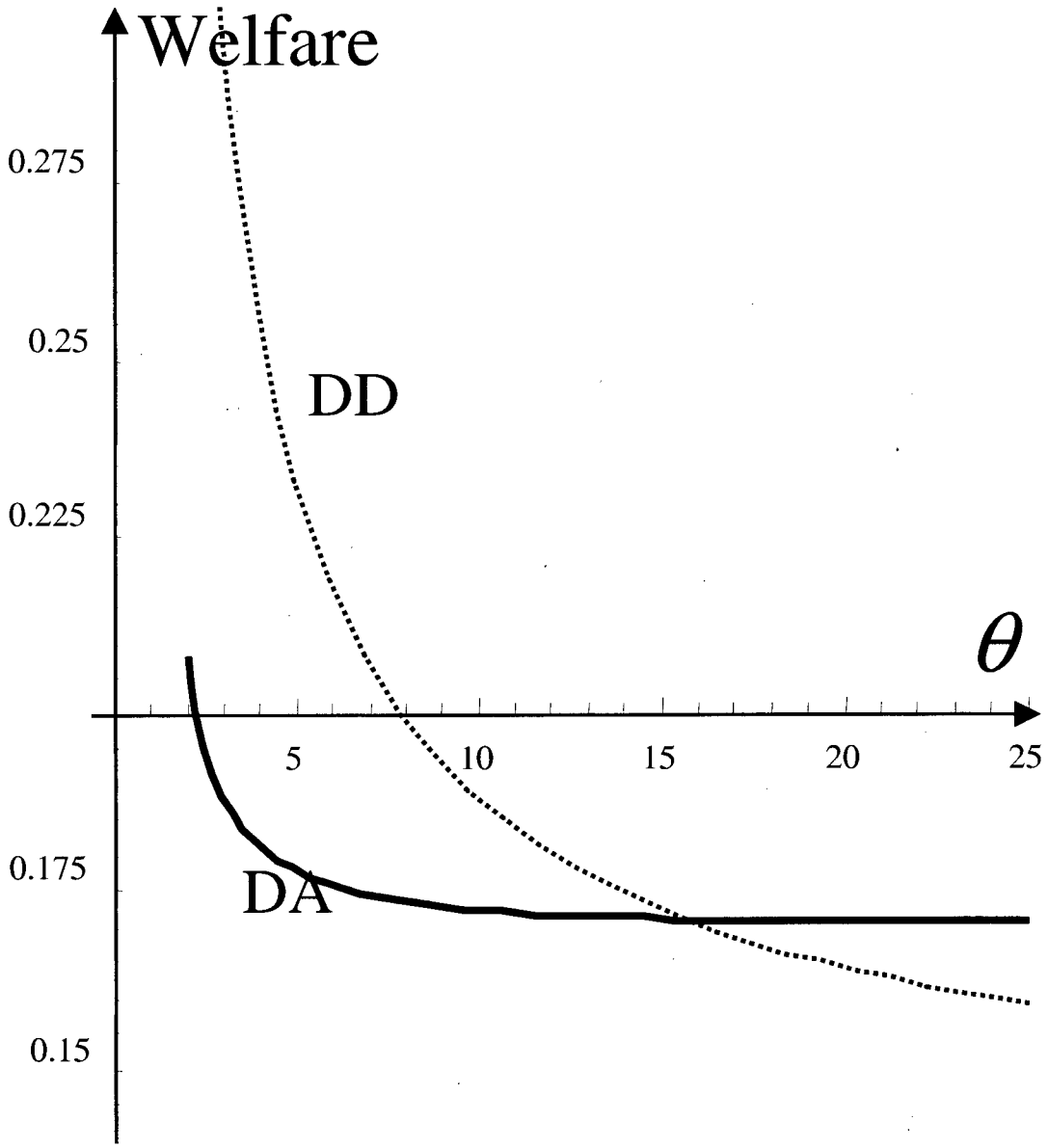

FIGURE 29. Welfare comparison between regimes DD and DA around the point $(\widetilde{F}^c, \widehat{F}^c)$

Parameters: $a = 1$, $c_0 = 0.5$

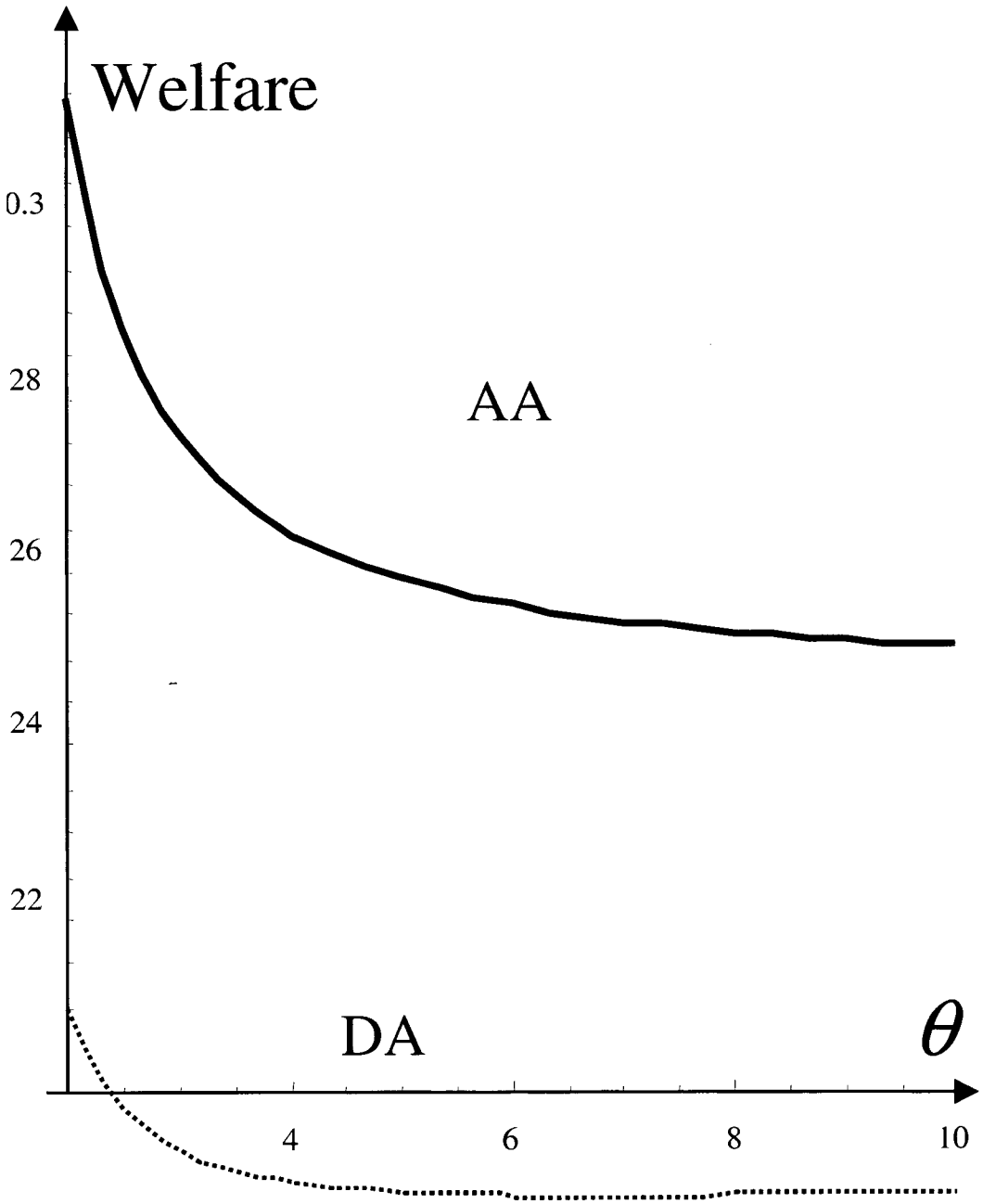

FIGURE 30. Welfare comparison between regimes AA and DA

Parameters: $a = 1$, $c_0 = 0.5$

CHAPTER 5

Conclusions

This dissertation aims to contribute to the understanding of the strategic implication of three main scenarios: commitments that can be revoked at a non-prohibitive cost, competition in the presence of capacity constraints and demand-independent markets that are connected by the fact there is only one firm competing in both of them with local firms. The results suggest these elements could have a large impact on market structure and performance.

In Chapter 2, I show that the effect of partial commitments differs drastically according to the type of game firms play. In the case of quantity competition, both firms can be hurt because the ability to commit intensifies competition and increase the total welfare generated by the industry. In the case of price competition with differentiated products, however, both firms can benefit because flexible commitments facilitate collusion and decrease total welfare. The stronger the commitments, the larger is their impact on welfare.

If commitment strength depends on the production technology, then it has to be taken into account in technology adoption decisions. In a quantity game, I show that there is an 'inflexibility trap': for strategic reasons, firms like the better commitment ability and in equilibrium they both can end up adopting the inflexible but inefficient technology, in spite of that fact that this intensifies competition and reduces their profits. The second mover always chooses a technology that is at least as flexible as the one adopted by the first mover.

The most surprising result with partial commitments is the fact that equilibrium prices can actually *increase* if a more efficient technology is adopted. This is due to the fact that the more efficient technology is associated with a lower ability to commit and this dilutes competition to the extent that the total welfare decrease.

Chapter 3 provides support for the traditional view about the effect of excess capacity on collusion possibilities. Using a tacit agreement in which firms produce the same output regardless of the ex-post installed capacity, I show that in the subgame perfect equilibrium firms choose a capacity lower than the one that minimizes the cost of producing the equilibrium output. This agreement removes the rent seeking behaviour characterizing the usual type of tacit agreement

used in the literature. Firms signal their willingness to collude by curtailing their capacities (increasing capacity does not bring along any strategic advantage).

The surprising result here is that when some exogenous changes help firms to collude (for example, a decrease increase in the number of firms), then the strategic role played by capacity in supporting collusion diminishes, and firms become more efficient. On one hand, an exogenous entry in the market can cause firms to become less efficient by further reducing their capacities but, on the other hand, the increased number of competitors tends to increase competition. It is possible that the decrease in efficiency overwhelms the intensified competition effect, thus leading to a decrease in welfare. An entry promoting policy cannot be guaranteed to be welfare improving if it does not attract a sufficiently large number of entrants in the market. Similarly, an exogenous increase in the discount factor makes firms more efficient and therefore can increase the per period welfare, in spite of producers now supporting a higher collusive price.

Finally, Chapter 4 analyzes interaction between two demand-independent market, indirectly linked by a national firm that competes with local firms in both markets. Entry in one market changes the profitability of R&D activities undertaken by the national firm to reduce its marginal cost of production. Depending on the R&D technology, the national firm becomes a tougher or softer competitor as a result and this affects the profit of firms and the market structure of the other market. If the national firm finds it profitable to intensify its R&D activities, prices decline in both markets, but the decline will be more dramatic in the entry market.

Surprisingly, if the regional markets have different entry costs and the national firm deters entry only in the high cost market, a policy aiming to encourage entry in that market by offering subsidies could yield a very different result: the national firm, seeing that the high cost market is now harder to exploit, might decide instead to deter entry in the other market as well. In the case in which the national firms deters entry in both markets, a subsidy in the *low* cost market might induce entry in *both* markets.

Bibliography

Amir, R. & Grilo, I. (1999), 'Stackelberg versus Cournot equilibrium', *Games and Economic Behavior* **26**, 1–21.

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

Brander, J. A. & Harris, R. G. (1984), Anticipated collusion and excess capacity. mimeo, UBC.

Brander, J. A. & Spencer, B. J. (1983), 'Strategic commitment with R&D: The symmetric case', *The Bell Journal of Economics* **14**(1), 225–235.

Brock, W. A. & Scheinkman, J. A. (1985), 'Price setting supergames with capacity constraints', *Review of Economic Studies* **52**(3), 371–382.

Bulow, J. I., Geankopolos, J. D. & Klemperer, P. D. (1985), 'Multimarket oligopoly: Strategic substitutes and complements', *Journal of Political Economy* **93**(3), 488–511.

Caruana, G. & Einav, L. (2002), A theory of endogenous commitment. Working Paper No. 0206, CEMFI.

Chen, Z. & Ross, T. W. (2002), Markets linked by rising marginal costs: Implications for multimarket contact, recoupment and retaliatory entry. mimeo.

Conlin, M. & Kadiyali, V. (1999), Capacity and collusion: an empirical analysis of the Texas lodging industry. mimeo, Cornell University.

Davidson, C. & Deneckere, R. (1990), 'Excess capacity and collusion', *International Economic Review* **31**, 521–542.

Davidson, P. (1963), 'Public policy problems of the domestic crude oil industry', *American Economic Review* **53**, 85–108.

Delbono, F. & Denicolo, V. (1991), 'Incentives to innovate in a Cournot oligopoly', *Quarterly Journal of Economics* **106**(3), 951–961.

Dixit, A. (1980), 'The role of investment in entry-deterrence', *Economic Journal* **90**(357), 95–106.

Eaton, B. C. & Eswaran, M. (1997), 'Supporting collusion by choice of inferior technologies', *Trade, technology and economics: Essays in honour of Richard G. Lipsey.* **Eaton,-B.-Curtis; Harris,-Richard-G. eds.**, 263–283.

Green, E. & Porter, R. (1984), 'Noncooperative collusion under imperfect price information', *Econometrica* **52**, 87–100.

Hamilton, J. & Slutsky, S. (1990), 'Endogenous timing in duopoly games: Stackelberg or Cournot equilibria', *Games and Economic Behavior* **2**, 29–46.

Henkel, J. (2002), 'The 1.5th mover advantage', *RAND Journal of Economics* **33**(1), 156–170.

McArthur, K. (2000), 'Air Canada accused of predatory pricing', *The Globe and Mail* p. B3.

Milgrom, P. & Roberts, J. (1982), 'Limit pricing and entry under incomplete information: An equilibrium analysis', *Econometrica* **50**, 443–460.

Nocke, V. (1999), Cartel stability under capacity constraints: The traditional view restored. (LSE) Economics of Industry Group Discussion Paper: EI/23.

Osborne, M. & Pitchik, C. (1987), 'Cartels, profits and excess capacity', *International Economic Review* **28**, 413–428.

Phlips, L. (1995*a*), *Competition policy: a game-theoretic perspective*, Cambridge University Press, New York.

Phlips, L. (1995*b*), *Excess capacity and collusion*, in , pp. 151–172.

Scherer, F. & Ross, D. (1990), *Industrial Market Structure and Economic Performance*, third edn, Houghton Mifflin Company.

Staiger, R. W. & Wolak, F. A. (1992), 'Collusive pricing with capacity constraints in the presence of demand uncertainty', *Rand Journal of Economics* **23**(2), 203–220.

Tirole, J. (1988), *The Theory of Industrial Organization*, The MIT Press.

Ware, R. (1984), 'Sunk costs and strategic commitment: A proposed three-stage equilibrium', *Economic Journal* **94**, 370–378.