14 Cartels and tacit collusion

Collusive (or price-fixing) agreements, whereby firms in an industry avoid competing with one another, play an important role in antitrust analysis. In this chapter, we focus on two important aspects of collusion. First, we shed some light on the incentives of firms to stay inside or outside an *explicit cartel*; that is, we focus on cartel formation. Here, cartel members are assumed to take their decisions jointly. For such explicit cartels to work, firms must enter into (long-term) binding agreements so as to form a joint profit-maximizing entity. The main issues here are the formation and the stability of such cartels: what is their optimal and equilibrium size? We study these issues in Section 14.1.

Second, we analyse how cartels and other forms of collusion can be sustained in the absence of binding agreements. Here, collusion emerges as the non-cooperative equilibrium of a situation of repeated competition. Because firms non-cooperatively adopt strategies that lead to a coordinated outcome, collusion is said to be *tacit*.[b] Since cartels are illegal, such tacit collusion is of particular importance for firms belonging to a cartel. For collusion to work, firms must have a correct 'understanding' of how other colluding firms will react to their behaviour. Therefore, the main issue in this second approach is the sustainability of tacit collusion: what is the set of prices that can be sustained in a non-cooperative equilibrium when competition is repeated over time? We examine this issue in Section 14.2.

There is a consensus that collusive agreements are welfare-reducing and should therefore be forbidden.[c] Note that price-fixing cartels stand out as one of the few areas in antitrust policy in which the inflexible 'per se' approach has not been challenged and replaced by a 'rule-of-reason' approach (as is the case, e.g., for the vertical restrictions examined in Chapter 17). As there is an agreement about what to do, the main issue is how to do it: how can antitrust authorities detect collusion and how should they fight it? This is what we study in Section 14.3.

14.1 Formation and stability of cartels

Although most competition laws forbid price-fixing agreements, explicit cartels continue to form and operate in a vast array of industries. As illustrated in Case 14.1 (to which we will return

[b] Tacit comes from the Latin *tacitus*, past participle of *tacere*, which means *'to be silent'*.

[c] It is nevertheless possible to provide arguments according to which price-fixing (perhaps not all the way up to monopoly level, but above the non-cooperative solution) may actually be *beneficial* for society. One argument is that the non-cooperative solution sustains too few firms and thus too little product variety in the market. Collusion, by increasing price–cost margins, leads to higher industry profits and, in effect, more entry. This additional entry may be socially desirable to ameliorate the previous social underprovision of the number of products. In addition, if a cartel not only shares information but in addition relocates capacities to more efficient use (which will often involve side-payments), overall production costs are reduced, which tends to raise welfare. In spite of these arguments, the general idea is that collusive agreements in the marketplace are bad for society.

several times throughout this chapter), cartels may last and maintain discipline over a long period of time. In this section we discuss cartels with explicit agreements and we address the issue of cartel formation and stability. This is important because in many markets, not all participants on the seller side collude, but only a subset of firms may form a cartel. In that case, we examine how such a cartel forms and what conditions keep a cartel member's incentives to stay in this group.

Case 14.1 The vitamin cartels

The worldwide market for bulk vitamins is estimated to be worth EUR 3.25 billion as of 1999. In Europe, sales of bulk vitamins were EUR 800 million in 1998 (with vitamins E, A and C accounting respectively for about EUR 250, 150 and 120 million). Production of vitamins is highly concentrated: the largest firm, Hoffmann-La Roche, has a market share of 40–50%, BASF has 20–30% and Aventis 5–15%. The three largest firms combined thus control between 65% and 95% of this market.[6] While slow and costly plant construction, as well as economies of scale in the production technology, foster concentration on the production side, the buyer side is more fragmented as bulk vitamins are bought by a large number of different companies, for example, producers of various food products, pharmaceutical companies and animal feed manufacturers. However, some buyers are likely to be large enough to enjoy some bargaining power in the form of lower prices.

On 21 November 2001, the European Commission imposed a then-record fine of EUR 855.22 million on eight companies for participating in secret market-sharing and price-fixing cartels affecting vitamin products.[7] Overall, the Commission assessed the existence of eight distinct cartels involving different vitamins between September 1989 and February 1999. This European case followed an investigation by the US Department of Justice, which in 1999 had already led to Hoffmann-La Roche pleading guilty and paying a US$500 million criminal fine for leading a worldwide conspiracy to raise and fix prices and allocate market shares for certain vitamins sold in the USA and elsewhere.[d] In this context, BASF also pleaded guilty and paid US$225 million in fines. In addition, there were large payments in civil settlements.

The European Commission found that overall 13 European and non-European companies participated in the cartels. The three main companies involved were Hoffmann-La Roche in Switzerland, BASF in Germany and Takeda Chemical Industries in Japan.[e] Overall the cartels involved the vitamins A, E, B1, B2, B5, B6, C, D3, biotin (H), folic acid (M), beta carotene and carotinoids. The first cartel arrangements were formed in the vitamin A and E markets, where they were also maintained for the longest time, ending in February 1999.

[6] See European Commission (2001a).

[7] See European Commission (2001b).

[d] The fine represented the largest fine that the US Department of Justice obtained in a criminal case up until then (see US Department of Justice, 1999).

[e] The other ten companies involved were Aventis in France (formerly Rhône-Poulenc), Daiichi Pharmaceutical in Japan, Eisai in Japan, Merck in Germany, Kongo Chemical in Japan, Lonza in Germany, Solvay Pharmaceuticals in the Netherlands, Sumitomo Chemical in Japan, Sumika Fine Chemicals in Japan and Tanabe Saiyaku in Japan.

Hoffmann-La Roche was active in every vitamin market concerned and BASF in all but two markets.

According to the European Competition Commissioner Mario Monti, the vitamin cartel was 'the most damaging series of cartels the Commission has ever investigated due to the sheer range of vitamins covered, which are found in a multitude of products from cereals, biscuits and drinks to animal feed, pharmaceuticals and cosmetics'. He concluded that 'the companies' collusive behaviour enabled them to charge higher prices than if the full forces of competition had been at play, damaging consumers and allowing the companies to pocket illicit profits'.[8]

When firms within an industry form a cartel, they eliminate the competition that was existing between them. This leads them to reduce their joint output or to increase their prices. As this collusive behaviour also benefits the firms outside the cartel, the formation of a cartel can be seen as a public good. One can therefore conjecture that firms will tend to free-ride on the cartels formed by other firms, making cartels highly unstable. However, as we now show, the stability of a cartel depends on the institutional procedures of group and network formation that govern the formation of this cartel.

We contrast three procedures.[9] In the first procedure, firms decide simultaneously whether or not to participate in a single industry-wide cartel; it is implicitly assumed that cartel membership is open, in the sense that firms cannot exclude other firms from the cartel. In contrast, the second procedure allows for the endogenous formation of multiple cartels in a sequential way and with exclusive membership; here, each firm anticipates how its decision will affect the behaviour of firms choosing actions subsequently in the game. Finally, the third procedure relies on bilateral market-sharing agreements, whereby pairs of firms refrain from competing on each other's territory; the collection of these bilateral agreements then constitutes a collusive network.

We compare the three procedures in an industry described by the following simple market structure: n symmetric firms produce a homogeneous good at constant marginal cost c; they compete à la Cournot and face an inverse demand given by $p = a - q$, where q is the total quantity produced. We will also indicate how the results may change under different market structures (e.g., when products are differentiated).

14.1.1 Simultaneous cartel formation

We start by considering a very simple game in which firms decide simultaneously whether or not they want to join a single cartel. If a firm joins the cartel, it jointly chooses with the other cartel members the quantity that maximizes their joint profits.[10] Otherwise, the firm remains independent and chooses the quantity that maximizes its own profits. Suppose that a cartel of k firms, with $1 < k \leq n$, is formed. The Cournot game is thus played among $(n - k)$ independent firms and the cartel made up of the other k firms. As all $(n - k + 1)$ players are symmetric (they have the same constant marginal cost c and face the same demand), we recall from

[8] See European Commission (2001b, p. 1).
[9] We follow here Bloch (2002).
[10] Note that these models can be seen as merger games in which firms can disintegrate.

Chapter 3 that each player gets profits equal to $[(a-c)/(n-k+2)]^2$ at the Cournot equilibrium. Because of symmetry, we assume that inside the cartel, the division of profits is equitable. Hence, for a given cartel size k, profits for firms inside and outside the cartel are respectively given by

$$\pi^{in}(k) = \frac{(a-c)^2}{k(n-k+2)^2} \quad \text{and} \quad \pi^{out}(k) = \frac{(a-c)^2}{(n-k+2)^2}.$$

Now, for the cartel to be stable, it must be that no cartel member has an incentive to unilaterally leave the cartel, which is equivalent to

$$\pi^{in}(k) \geq \pi^{out}(k-1) \Leftrightarrow \frac{(a-c)^2}{k(n-k+2)^2} \geq \frac{(a-c)^2}{(n-k+3)^2}.$$

The latter condition can be rewritten as follows:

$$\begin{aligned} &(n-k+3)^2 \geq k(n-k+2)^2 \\ \Leftrightarrow\ &(n-k)^2 + 6(n-k) + 9 \geq k(n-k)^2 + 4k(n-k) + 4k \\ \Leftrightarrow\ &(1-k)(n-k)^2 + (6-4k)(n-k) + (9-4k) \geq 0. \end{aligned}$$

It is easily seen that for $k \geq 3$, which supposes $n \geq 3$, all the terms of the above inequality are negative. This implies that the condition is always violated, meaning that whatever the size of the cartel, each member has an incentive to leave it. For $k = 2$, the condition becomes $-n^2 + 2n + 1 > 0$, which is never satisfied for $n > 2$. If $n = 2$, the only possible cartel comprises all firms (i.e., $k = n = 2$) and the above inequality is then satisfied. We have thus proved the following result.

Lesson 14.1 **Consider the formation of a single cartel on a Cournot market with homogeneous goods and constant marginal costs. If there are at least three firms in the industry, all firms remain independent. If there are just two firms in the industry, the two firms form a cartel.**

The intuition for this result is simple: because the formation of the cartel induces positive externalities on the firms outside the cartel (through the higher market price), all firms prefer to free-ride on the public good provided by cartel members. In the linear Cournot model with constant marginal costs, this free-riding incentive is so strong that it prevents the formation of any cartel.

However, when firms produce horizontally differentiated goods, competition is relaxed and so is the free-riding incentive.[f] To examine the effects of product differentiation on cartel stability, suppose that each of the n firms in the industry produces a differentiated variety at a constant marginal cost c. Consider the following inverse demand functions $p_i = a - q_i - \gamma \sum_{j \neq i} q_j$, where $\gamma \in [0,1]$ measures the strength of product substitutability.[g]

Two important results emerge from the analysis of the cartel formation game in this model. First, it is possible to show that, unless $n = 2$, the full cartel is not stable, whatever

[f] To be sure, in the extreme case of perfectly differentiated goods, cartelization would make no difference as firms would be monopolies; firms would therefore be indifferent between being inside or outside the cartel.

[g] As shown in Chapter 3, these inverse demand functions are derived from the maximization of a quadratic utility function of a representative consumer having a taste for variety.

the degree of product substitutability – the reader may want to check this. There is always an incentive for an individual firm to free-ride on the cartel comprising all the other firms: for $n \geq 3$, $\pi^{out}(n-1) > \pi^{in}(n)$ for all $\gamma \in (0, 1]$.

Second, partial cartels may be stable, as illustrated by the following example. Let $n = 3$, $\gamma = 1/2$ and $a - c = 1$. We want to show that the cartel formation game has a Nash equilibrium in which two firms form a cartel while the third firm remains independent. In this situation, firms in the cartel (say firms 1 and 2) choose quantities q_1 and q_2 so as to maximize their joint profits $\Pi_{12} = (1 - q_1 - \frac{1}{2}(q_2 + q_3))q_1 + (1 - q_2 - \frac{1}{2}(q_1 + q_3))q_2$. As the two firms are symmetric, the maximum is such that $q_1 = q_2 \equiv q_{12}$, and the joint profits can be rewritten as $\Pi_{12} = 2(1 - \frac{3}{2}q_{12} - \frac{1}{2}q_3)q_{12}$. From the first-order condition, we derive the cartel's reaction function as $q_{12}(q_3) = \frac{1}{6}(2 - q_3)$. The independent firm chooses q_3 to maximize $\pi_3 = (1 - q_3 - \frac{1}{2}(q_{12} + q_{12}))q_3$. Again, we obtain the reaction function from the first-order condition: $q_3(q_{12}) = \frac{1}{2}(1 - q_{12})$. Solving for the system of the two reaction functions, we find the equilibrium quantities $q_{12} = 3/11$ and $q_3 = 4/11$. Computing the equilibrium profits, we have $\pi_1 = \pi_2 = \pi^{in}(2) = 27/242$ and $\pi_3 = \pi^{out}(2) = 32/242$. Observe that the free-riding effect is still present as the outsider obtains a larger profit than the insiders.

There are two requirements for the cartel to be stable: first, no cartel member has an incentive to leave the cartel; second, no outside firm finds it profitable to join the cartel. The first requirement refers to the *internal stability* of the cartel, and the second to its *external stability*.[11] Checking first for internal stability, we must compute the profits a firm would obtain by leaving the cartel. In that case, the three firms would be independent and each firm would choose its quantity q_i to maximize $\pi_i = (1 - q_i - \frac{1}{2}(q_j + q_k))q_i$. The first-order condition gives $1 - 2q_i - \frac{1}{2}(q_j + q_k) = 0$. Invoking symmetry, each firm produces the same quantity q at the equilibrium and the latter condition becomes: $1 - 3q = 0$. Hence, $q = 1/3$ and equilibrium profits are $\pi^{out}(1) = 1/9$. Internal stability is fulfilled as long as $\pi^{in}(2) \geq \pi^{out}(1)$, which is true as $27/242 \simeq 0.1116 > 1/9 \simeq 0.1111$.

To check for external stability, we must compute the profits firm 3 would obtain by joining firms 1 and 2 in the cartel. In that case of full collusion, the three firms choose their common quantity q so as to maximize $\Pi_{123} = 3(1 - q - \frac{1}{2}(2q))q = 3(1 - 2q)q$. The profit-maximizing quantity is easily found as $q = 1/4$, which results in per-firm profits of $\pi^{in}(3) = 1/8$. Hence, there is no incentive for the outside firm to join the cartel if $\pi^{out}(2) \geq \pi^{in}(3)$, which is satisfied as $32/242 \simeq 0.132 > 1/8 = 0.125$.

When products are sufficiently differentiated, competition is less intense and, as a result, outside firms do not benefit so much from the formation of a cartel. In other words, the free-riding incentive is reduced with respect to the case of homogeneous goods. In our example, the reduction is strong enough to dissuade one of the two cartel members to leave the cartel; in contrast, free-riding is still attractive enough to dissuade the outside firm from joining the cartel.

Lesson 14.2 **Consider the formation of a single cartel on a Cournot market with differentiated goods. If goods are sufficiently differentiated, it is possible to find stable cartels comprising not all firms but a strict subset of them.**

---

[11] This typology is due to d'Aspremont *et al.* (1983).

14.1.2 Sequential cartel formation

We now allow for the formation of multiple cartels. In such a case, the profit of each firm will depend on the entire cartel structure. Hence, each firm will have to anticipate the other firms' reaction when deciding whether to join a cartel or not. To take this forward-looking behaviour into account, we assume here that the production stage (i.e., the Cournot game) is preceded by a sequential game of cartel formation.[12] This sequential game is defined by an exogenous specification of the ordering of the firms that determines the order of move. It proceeds as follows. The first firm in the ordering proposes the formation of a cartel (to which it belongs) and all the prospective members of this cartel sequentially respond in turn to the proposal (note that the sequence of all firms is initially determined according to some exogenous rule). If they all agree, the proposed cartel is formed in the order initially determined. If at least one firm rejects the first proposal, the cartel is not formed and it is the first firm that rejected the offer that makes a counteroffer and proposes a cartel to which it belongs. If all firms accept the first proposal, the cartel is formed and these firms withdraw from the game. The first among the remaining firms then makes a proposal and the game proceeds. To make sure that cartels will eventually be formed (with, possibly, all firms remaining separate) and the Cournot game will ensue, it is assumed that there is no discounting in the cartel formation game but that firms receive zero profits if the game is played infinitely.[h]

The game has complete information and infinite horizon; we therefore use the solution concept of a *stationary perfect equilibrium*.[i] Note the differences with the previous setting: cartel formation is now sequential (rather than simultaneous) and membership is exclusive (rather than open). Moreover, as the game is only played among the remaining firms once a cartel has been formed, this sequential cartel formation embodies a high degree of commitment of the firms.

The stationary perfect equilibrium of this game has a very simple structure: at each period *the first firms in the sequence choose to remain independent and to free-ride on the cartel that the last firms will eventually form.* To find the critical cartel size, we need to compare the profits a firm obtains in a cartel of a given size, $\pi^{in}(k)$, with the profits each firm obtains when they all remain independent, $\pi^{out}(1)$. Firms prefer to form a cartel than to remain independent if

$$\pi^{in}(k) \geq \pi^{out}(1) \Leftrightarrow \frac{(a-c)^2}{k(n-k+2)^2} \geq \frac{(a-c)^2}{(n+1)^2}.$$

Developing the latter condition, we have

$$(n+1)^2 \geq k(n-k+2)^2 \Leftrightarrow (k-1)\left(-k^2+(2n+3)k-(n+1)^2\right) \geq 0.$$

[12] This model is proposed by Bloch (1996).

[h] When firms are ex ante identical (as is assumed here), Bloch (1996) shows that a simpler finite game generates the exact same cartel structures as those obtained at the stationary perfect equilibria of the infinite game. In this finite game, the first firm announces an integer k_1, corresponding to the size of the cartel it wants to see formed, firm k_1+1 announces an integer k_2 and so on until the total number n of firms is exhausted. An equilibrium of the finite game determines a sequence of integers adding up to n, which characterizes the cartel structure completely as all firms are ex ante identical.

[i] Stationarity requires that strategies only depend on the payoff-relevant part of the history (i.e., here, the cartel structure formed by the previous players and the ongoing offer).

Solving for k, we have that[13]

$$\pi^{in}(k) > \pi^{out}(1) \Leftrightarrow k > \tfrac{1}{2}\left(2n + 3 - \sqrt{4n+5}\right).$$

Let k^* denote the first integer following $\frac{1}{2}(2n + 3 - \sqrt{4n+5})$. Then, the first $n - k^*$ firms indeed prefer to remain independent. However, the last k^* firms will choose to form a cartel as it is no longer profitable for a firm to stay out and free-ride on the cartel formed by the others. In this specification we show that the minimal profitable cartel size is larger than 80% of the firms in the industry:

$$k^* - \frac{80}{100}n = \frac{1}{10}\left(2n + 15 - 5\sqrt{4n+5}\right) = \frac{1}{10}\frac{(2n+15)^2 - 25(4n+5)}{2n + 15 + 5\sqrt{4n+5}}$$

$$= \frac{1}{10}\frac{4(n-5)^2}{2n + 15 + 5\sqrt{4n+5}} \geq 0.$$

It follows that there will always be at most one cartel formed at equilibrium.

Lesson 14.3 **Consider a Cournot market with homogeneous goods. In the unique equilibrium of the sequential cartel formation game, the first $(n - k^*)$ firms remain independent while the last k^* firms form a cartel, with k^* being larger than 80% of the firms in the industry.**

Recall that in the simultaneous cartel formation game, any firm has an incentive to leave the cartel. Here, the sequentiality allows firms to commit to stay out of the cartel, which makes it possible for a cartel to form at equilibrium. In that respect, the cartel formed in this sequential way has the attributes of a complete merger of the firms. A horizontal merger can indeed be seen as a cartel requiring the unanimity of its members: if one member disagrees, the cartel does not form (i.e., the merger is not implemented). We will show in the next chapter that the '80% rule' stated in Lesson 14.3 also applies to horizontal mergers of symmetric firms in a Cournot industry.

14.1.3 Network of market-sharing agreements

In the previous two situations, collusion resulted from the formation of multilateral agreements. Here, we instead consider *bilateral* collusive agreements. In particular, we talk about *market-sharing agreements* of the following form: if two firms are active on different geographical markets or serve distinct consumer segments, they may collude by signing a market-sharing agreement, whereby they both refrain from competing on the other firm's territory (we return to such agreements, and give illustrations, in the next section). At the industry level, the collection of such bilateral market-sharing agreements constitutes a collusive structure.

We refer to this collusive structure as a *collusive network.* A network is represented by a graph g on the set N of firms. A graph is a set of pairwise links, denoted ij, between firms i and j. Here, the link ij is formed if firms i and j sign a market-sharing agreement. The larger the number of firms, the more complex are the potential networks. It is therefore very hard to characterize the

[13] The polynomial in the second bracket has two roots: $(1/2)(2n + 3 \pm v(4n + 5))$. The larger root is clearly larger than n, while the smaller root can be shown to be positive.

formation of a network as the result of a non-cooperative game of link formation. Hence, we just focus here on the issue of network stability. We say that network g is stable if no pair of firms has an incentive to form a new link, and no firm has an incentive to unilaterally destroy an existing link.[14] Applied to the present situation, a collusive network is stable if no pair of firms finds it profitable to sign a new market-sharing agreement, and no firm has an incentive to renege on an existing market-sharing agreement.

Suppose that the n firms are initially present on different geographical markets and that each market is characterized by Cournot competition over homogeneous goods. As above, we assume that the inverse demand (on each market) is given by $p = a - q$, and that all firms have the same constant marginal cost of production c. For a given collusive network g, let $n_i(g)$ denote the number of firms active on market i. The total profit of firm i (initially installed on market i) can then be written as

$$\Pi_i(g) = \frac{(a-c)^2}{(n_i(g)+1)^2} + \sum_{j \text{ with } ij \notin g} \frac{(a-c)^2}{(n_j(g)+1)^2},$$

where the first term is the Cournot equilibrium profit that firm i obtains in its home market, and the second term is the sum of profits firm i obtains in all the foreign markets in which it is present (i.e., the markets of the firms with which it has not signed a market-sharing agreement). Denoting by $g + ij$ the graph obtained by adding link ij to graph g, we can use the previous definition to compute firm i's *incentive to form an agreement* with firm j as

$$\Pi_i(g+ij) - \Pi_i(g) = \underbrace{\left[\frac{(a-c)^2}{n_i(g)^2} - \frac{(a-c)^2}{(n_i(g)+1)^2}\right]}_{\text{less competition on } i\text{'s market}} - \underbrace{\frac{(a-c)^2}{(n_j(g)+1)^2}}_{\text{lost access to } j\text{'s market}}. \tag{14.1}$$

There are two conflicting effects at work when forming a new agreement. On the one hand, firm i benefits from the reduction of competition on its own market (as firm j has withdrawn); on the other hand, firm i foregoes access to market j and the profits it was making there. (These are the only two effects as the other markets $k \neq i, j$ are left unaffected by the market-sharing agreement between firms i and j.)

We first examine the *stability of the empty network*, that is, the situation in which no market-sharing agreement is signed and no collusion occurs. In the absence of any agreement, all n firms are present on all markets. Hence, using expression (14.1) and writing g_0 for the empty network, we compute the benefit that two firms would have to sign an agreement as

$$\Pi_i(g_0+ij) - \Pi_i(g_0) = \frac{(a-c)^2}{n^2} - 2\frac{(a-c)^2}{(n+1)^2} = \frac{(a-c)^2(-n^2+2n+1)}{n^2(n+1)^2}.$$

It is quickly seen that $(-n^2+2n+1)$ is positive for $n = 2$ and negative for $n \geq 3$. It follows that for $n \geq 3$, no pair of firms has an incentive to form a new link, meaning that the empty network is stable for $n \geq 3$.

[14] This notion of pairwise stability has been introduced by Jackson and Wolinsky (1996). The analysis of networks of bilateral market-sharing agreements is due to Belleflamme and Bloch (2004).

Suppose now that two firms in a non-empty network g have an incentive to sign an agreement: $\Pi_i(g) - \Pi_i(g - ij) \geq 0$ and $\Pi_j(g) - \Pi_j(g - ij) \geq 0$, or

$$\begin{cases} \frac{(a-c)^2}{(n_i(g)+1)^2} \geq \frac{(a-c)^2}{(n_i(g)+2)^2} + \frac{(a-c)^2}{(n_j(g)+2)^2}, \\ \frac{(a-c)^2}{(n_j(g)+1)^2} \geq \frac{(a-c)^2}{(n_j(g)+2)^2} + \frac{(a-c)^2}{(n_i(g)+2)^2}. \end{cases}$$

We note first that because of symmetry, this system of inequalities can only be satisfied if $n_i(g) = n_j(g)$: a market-sharing agreement can only be concluded among two firms with the same number of competitors on their home markets. Indeed, if one market had a smaller number of competitors, the firm on the other market would have no incentive to form an agreement, as the profit it makes on the foreign market would already be larger than the profit it makes on its home market. Second, setting $n_i(g) = n_j(g) = x$, the two inequalities can be rewritten as

$$\frac{(a-c)^2}{(x+1)^2} \geq 2\frac{(a-c)^2}{(x+2)^2} \Leftrightarrow (x+2)^2 \geq 2(x+1)^2 \Leftrightarrow 2 - x^2 \geq 0.$$

As $x = 1$ is the only strictly positive number that satisfies the latter inequality, it follows that if a non-empty network is formed, it must be such that each firm is a monopoly on its home market ($n_i(g) = 1$). In other words, *the only non-empty stable network is the complete network*, in which all pairs of firms sign market-sharing agreements, resulting in full collusion. Any firm would indeed have an incentive to defect from any network of smaller size.

Lesson 14.4 **If collusive networks are negotiated bilaterally, they may lead to full collusion, with every firm a monopolist on its own market.**

As the empty network is also stable (except for $n = 2$), we see that the free-riding incentives remain strong. However, they do not prevent the formation of full collusion. As above, when products are differentiated, competition is reduced and with it, the incentives to sign market-sharing agreements. Different stable network architectures might then emerge (e.g., a network with one isolated firm and all other $(n - 1)$ firms signing agreements with one another).

14.2 Sustainability of tacit collusion

In this section, we consider situations where collusion is reached in a 'tacit', non-cooperative, way. Firms do not necessarily resort, for example, to explicit quota agreements to discipline collusion. Instead, all that is needed is some 'meeting of the minds' between colluding firms and a common understanding that deviation from the collusive 'tacit agreement' will be met by some form of punishment. However, the analysis is also highly relevant for explicit agreements: sustainability is necessary for cartels that are reached in a cooperative way as long as punishments cannot be legally binding.

We start by analysing a simple repeated game framework to understand what this 'meeting of the minds' means exactly (Subsection 14.2.1). We then extend the basic setting in two directions: we first ask how punishments should optimally be designed to promote collusion and second how multimarket contact can make collusion easier to sustain (Subsections 14.2.2 and 14.2.3). We close this section with two more advanced topics, analysing how demand fluctuations and the unobservability of firms' actions impact the sustainability of tacit collusion (Subsections 14.2.4 and 14.2.5).

14.2.1 Tacit collusion: the basics

Consider a number of firms (say two, for simplicity) that offer perfect substitutes produced at constant marginal costs c. Suppose that instead of competing just once, the two firms repeatedly compete over time. That is, at each period $t = 1, 2, \ldots, T$, the firms repeat the 'static' game we analysed in Chapter 3 by simultaneously choosing the price, or the quantity, of their product. They observe the rival's action after each period. A strategy is then a list of actions (a contingent plan) that tells the firm what to choose in each period as a function of past prices or quantities (that is, the history of the game). Strategies can take complicated forms; in particular, the action prescribed at some period may depend on the observed actions (of all players) at all previous periods. Future periods are discounted by the factor δ and thus present discounted profits are $\sum_{t=0}^{T} \delta^t \pi_i^t(\sigma_1^t, \sigma_2^t)$, where π_i^t is firm i's profit in period t, σ_i^t is firm i's action in that period (with $\sigma_i = p_i$ or $\sigma_i = q_i$ according to whether we consider price or quantity competition) and T is the last period (the 'horizon') of the repeated game.

The question we want to investigate is whether firms can design strategies that allow them to reach a situation of tacit collusion, whereby they share the monopoly profit in each and every period. In what follows, we first show that whether the horizon is finite or infinite makes a crucial difference. We then compare price and quantity competition and examine the impact of the number of firms on the sustainability of tacit collusion.

Finite vs. infinite horizon

Suppose that the static game that firms play has a unique Nash equilibrium. We first want to show that *tacit collusion is not possible when competition is repeated over a finite number of periods* (that is, when competition stops at some finite date T). Since there is a known end date to the game, we can use backward induction to solve the game for its subgame-perfect equilibria. Consider thus the last period (period T). At this period, firms only care about their current profits as nothing ensues: the Nash equilibrium is thus that of the static game and each firm earns the level of profits at this equilibrium (denoted by π^n). Moving back to the penultimate period ($T - 1$), we now have that firms maximize their flow of profits over the last two periods. Yet, from our previous result, we know that whatever action they choose in period $T - 1$, this will not affect the profits they obtain in the subsequent period (i.e., π^n); therefore, the Nash equilibrium in period $T - 1$ is thus again the Nash equilibrium of the static game. The same argument carries on up to the first period. We have thus learned the following lesson.

Lesson 14.5 **If competition is repeated over a finite number of periods, firms play according to the (unique) Nash equilibrium of the static game in each period. Tacit collusion cannot emerge.**

Consider now an *infinite horizon* ($T = \infty$). This is not to say that firms compete until the end of time; it just means that there is no known end date to the game: at each period, there is a probability that firms will compete one more time. The first thing to note is that the subgame-perfect equilibrium we described for the finite horizon case is still an equilibrium here. Think of the strategy that tells the firm to choose, in each period and irrespective of what happened in the past, the action corresponding to the Nash equilibrium of the static game. Clearly, if firm i follows this strategy, it is firm j's best response to follow this strategy as well. The Nash equilibrium of the static game is thus repeated infinitely and no tacit collusion takes place. What is interesting is that this situation is no longer the only equilibrium. In particular, tacit collusion may emerge at the subgame-perfect equilibrium of the (infinitely) repeated game.

To see this, consider the so-called *grim trigger strategy*. According to this strategy, firm i starts by choosing the action that maximizes total profits (i.e., the sum of firm 1's and firm 2's profits). Firm i keeps on choosing this action as long as both firms have done so in all previous periods. We are then in the *cooperation phase*. However, if one firm deviates (i.e., chooses any other action), this deviation 'triggers' the start of the *punishment phase*: from the next period on, and forever after, both firms choose the action that corresponds to the Nash equilibrium of the static game.

We define the following per-period profit levels: when both firms play the cooperative action, they both obtain $\pi^c = \pi^m/2$ (where π^m is the per-period monopoly profit); when one firm plays the cooperative action and the other deviates optimally, the deviating firm obtains π^d; at the Nash equilibrium of the static game, both firms obtain π^n. Naturally, we have $\pi^d > \pi^c > \pi^n$. Since firms maximize their flow of profits over all the periods of the game, they will compare the immediate gain from deviation with future losses resulting from the other firm's punishment. As we will now show formally, this trade-off (and thus the feasibility of tacit collusion) depends (i) on the magnitude of the deviation and the punishment profits, with respect to the collusive profits, and (ii) on the firms' discount factor.

Since there is no terminal period to the game, we cannot solve it by backward induction. We proceed differently by supposing that one firm (say firm 2) follows the grim trigger strategy and by investigating under which conditions firm 1's best response is to follow the grim trigger strategy as well. Suppose first that we are in the punishment phase. Firm 2 then plays the action corresponding to the Nash equilibrium of the static game (now and forever) and, obviously, firm 1's best response is to play the same action at each period. Hence, firm 1 finds it best to follow the trigger strategy.[j] Consider now the cooperative phase. If firm 1 follows the grim trigger strategy, it will obtain π^c in all subsequent periods (as the punishment phase will never be triggered). In present discounted value, firm 1 will obtain

$$V^C = \pi^c + \delta\pi^c + \delta^2\pi^c + \cdots = \frac{\pi^c}{1-\delta}.$$

On the contrary, if firm 1 deviates, it will obtain π^d in the current period and π^n in all subsequent periods as the punishment phase will start. In present discounted value, deviation would give

$$V^D = \pi^d + \delta\pi^n + \delta^2\pi^n + \cdots = \pi^d + \delta\pi^n/(1-\delta).$$

[j] This makes the punishment credible. Hence, it can be played in subgame-perfect equilibrium off the equilibrium path.

Hence, firm 1 prefers to follow the grim trigger strategy if and only if

$$V^C \geq V^D \Leftrightarrow \frac{\pi^c}{1-\delta} \geq \pi^d + \frac{\delta}{1-\delta}\pi^n$$
$$\Leftrightarrow \frac{\delta}{1-\delta}(\pi^c - \pi^n) \geq \pi^d - \pi^c,$$

where the left-hand side is the discounted long-term loss induced by the punishment and the right-hand side is the short-term gain from deviation. Solving for δ, we can rewrite the previous condition as

$$\delta \geq \frac{\pi^d - \pi^c}{\pi^d - \pi^n} \equiv \delta_{\min}, \tag{14.2}$$

where $\delta_{\min}$ lies strictly between 0 and 1. We conclude the following.

Lesson 14.6 **When competition is repeated over an infinite horizon, tacit collusion can be sustained by the grim trigger strategy as long as firms have a large enough discount factor.**

The intuition is clear: firms must put sufficient weight on future losses to offset the temptation of securing an immediate gain by deviating; the minimum weight is lower (and thus tacit collusion is more likely) when deviation pays less (i.e., when the difference $\pi^d - \pi^c$ decreases) and punishment hurts more (i.e., when the difference $\pi^c - \pi^n$ increases). Note that this result hinges on the fact that firms are able to observe deviations from the collusive outcome and then start the punishment phase. In Subsection 14.2.5, we examine what changes when deviations are only imperfectly observable.

If condition (14.2) is met, then there exists a subgame-perfect equilibrium in which firms share the monopoly profit in each period. Note that there are plenty of other equilibria in the game. Actually, we can use the exact same argument to show that, if δ is large enough, any profit level $\tilde{\pi} \in [\pi^n, \pi^m/2]$ can be achieved by each firm in each period by setting the corresponding action $\tilde{\sigma}$ such that $\pi(\tilde{\sigma}, \tilde{\sigma}) = \tilde{\pi}$. Thus, this infinitely repeated game presents us with an embarrassment of riches; this result is commonly referred to in the literature as the *Folk theorem*.[15]

Application to price competition

Consider the simple Bertrand competition model with constant and identical marginal costs of production that we introduced in Chapter 4. If firms set the same price, the market is assumed to split evenly. Thus, if both firms collude and set the monopoly price p^m, they make in each period a profit of $\pi^c = \pi^m/2$. At any point in time, the largest short-term gain is achieved by slightly undercutting the rival's price by some arbitrarily small amount. Then deviation profits are $\pi^d = \pi^m - \varepsilon$. Since the total profit loss from setting a price slightly below p^m is negligible, we can drop ε in the analysis below. After a deviation has occurred, each firm sets the Nash equilibrium price c and makes zero profit: $\pi^n = 0$. Substituting for these values into the expression of the

15 See, for example, Friedman (1971, 1977).

minimum discount factor (14.2), we obtain the following:

$$\delta_{\min}^{\text{Bert}} = \frac{\pi^d - \pi^c}{\pi^d - \pi^n} = \frac{\pi^m - (\pi^m/2)}{\pi^m - 0} = \frac{1}{2}.$$

To illustrate the Folk theorem, note that if firms set $\widetilde{p} \in [c, p^m]$, then the total profit is $\pi(\widetilde{p}) = \widetilde{\pi}$, with $\widetilde{\pi} \in [0, \pi^m]$. In that case, the collusive profit is $\pi^c = \widetilde{\pi}/2$, deviation through undercutting yields $\pi^d = \widetilde{\pi}$, while in the punishment phase, each firm still obtains $\pi^n = 0$. Therefore, the minimum discount factor for sustaining collusion is the same as in the case where both firms set the monopoly price. We can thus state the following result.

Lesson 14.7 **In the infinitely repeated Bertrand price-setting duopoly game, any profit level between zero and the monopoly profit can be supported in a subgame-perfect equilibrium if the discount factor is sufficiently large, $\delta \geq 1/2$.**

Let us extend this simple example to the situation in which n firms operate in the market. Deviation profits remain unchanged ($\pi^d = \pi^m$) but total collusive profits have to be shared among n firms ($\pi^c = \pi^m/n$). Therefore, the condition that a deviation from the fully collusive outcome is not profitable becomes

$$V^C \geq V^D \Leftrightarrow \frac{1}{1-\delta}\frac{\pi^m}{n} \geq \pi^m \Leftrightarrow \delta \geq 1 - \frac{1}{n} \equiv \delta_{\min}^{\text{Bert}}(n).$$

The critical discount factor $\delta_{\min}^{\text{Bert}}(n)$ is increasing in n. This means that collusion becomes more difficult to sustain (in the sense that it requires less discounting) as the number of firms in the market increases. This suggests that collusion is more likely to be observed in markets in which a small number of firms are active.

Lesson 14.8 **In the infinitely repeated Bertrand price-setting game, the set of discount factors that can support collusion is larger the smaller the number of firms in the market.**

As has been pointed out above, though cartels rely on explicit agreements, firms cannot appeal to courts to enforce these agreements (as they are illegal) and must therefore rely on self-enforcing mechanisms. Thus, the previous finding applies equally to cartels, like the vitamin cartels that we described in Case 14.1.

Case 14.2 The vitamin cartels (2)

The sustainability of the vitamin cartels was fostered by three key factors. First, the production of vitamins is highly concentrated and associated with high entry costs, thus coordination among a small number of manufacturers could be effective and a strong firm could enforce existing agreements as well as threaten to punish deviating firms. Second, bulk vitamins are a homogeneous product with clear quality and product standards giving the transparency necessary to run a cartel successfully. Third, demand can be considered relatively inelastic, giving rise to a strong incentive to increase prices above the competitive level.

It has been observed that cartelization led to an increase in prices. Even more pronounced were price drops at the end of the cartelization period: it has been reported that price drops of 50% occurred for most cartelized vitamins shortly after announcement of the guilty pleas in the USA.[16] While prices are easily observable, the profitability and long-term effects of cartelization are more difficult to observe. In particular, demand and supply factors may have changed over time. Thus, it is unclear what the industry would have looked like in the absence of the cartel.[k]

Application to quantity competition

We now analyse the linear n-firm Cournot model with constant and identical marginal costs of production: the inverse demand is given by $P(q) = a - q$ and all firms have marginal costs of production c. The monopoly quantity is easily found as $q^m = (a-c)/2$, resulting in collusive profits

$$\pi^c = \pi^m/n = \frac{1}{4n}(a-c)^2. \tag{14.3}$$

We also recall from Chapter 3 that the Cournot Nash equilibrium profits in this linear model are

$$\pi^n = \frac{1}{(n+1)^2}(a-c)^2.$$

If all other firms choose a quantity q, the best deviation is the quantity z that maximizes $z(a-c-z-(n-1)q)$. From the first-order condition, we find $z = (1/2)(a-c-(n-1)q)$. Substituting for this value in the profit function, we obtain

$$\pi^d(q) = \tfrac{1}{4}(a-c-(n-1)q)^2. \tag{14.4}$$

Hence, if the other firms play the collusive quantity $q^m/n = (a-c)/(2n)$, we have

$$\pi^d = \frac{(n+1)^2}{16n^2}(a-c)^2.$$

We now have all the elements to compute the minimum discount factor that allows firms to sustain the monopoly outcome. Using expression (14.2), we obtain

$$\delta_{\min}^{\text{Cour}}(n) \equiv \frac{\pi^d - \pi^c}{\pi^d - \pi^n} = \frac{\frac{(n+1)^2(a-c)^2}{16n^2} - \frac{(a-c)^2}{4n}}{\frac{(n+1)^2(a-c)^2}{16n^2} - \frac{(a-c)^2}{(n+1)^2}} = \frac{(n+1)^2}{n^2+6n+1}.$$

It can be checked that $\delta_{\min}^{\text{Cour}}(n)$ increases with n. For instance, in a duopoly, the minimum discount factor is $9/17 \approx 0.53$; if there are three firms in the market, the value increases to $4/7 \approx 0.57$; for ten firms, we reach $121/161 \approx 0.75$, and so on and so forth. Therefore, as under price competition,

[16] For details on prices, see Connor (2001).

[k] In general, a more detailed understanding of the effects of cartelization on industry performance would be desirable. In their survey of the empirical analysis of cartels in a wide number of industries, Levenstein and Suslow (2006) conclude that cartels 'appear to increase prices and profits, but more careful studies with explicit counterfactual analysis would make a significant contribution to our understanding of the full economic effects of collusion'.

collusion is harder to sustain the larger the number of firms in the market. In the limit as n tends to infinity, the critical discount factor converges to 1.

Note that in this linear example, collusion is easier to sustain under price competition than under quantity competition when the market is a duopoly: $\delta_{\min}^{\text{Bert}}(2) = 0.5 < \delta_{\min}^{\text{Cour}}(2) \simeq 0.53$. However, the opposite result prevails when there are at least three firms in the market; one can indeed check that for $n \geq 3$, $\delta_{\min}^{\text{Bert}}(n) > \delta_{\min}^{\text{Cour}}(n)$. It is not surprising that the comparison between the two types of competition leads to ambiguous results insofar as both the gain from deviation $(\pi^d - \pi^c)$ and the impact of punishment $(\pi^c - \pi^n)$ are larger under price competition than under quantity competition. For $n = 2$, the stronger punishment makes collusion easier to sustain under price competition; for $n \geq 3$, the stronger incentive to deviate makes collusion harder to sustain under price competition.

To sum up, we have shown in this subsection that firms can sustain collusion in an infinitely repeated interaction provided the future is a sufficiently strong component of the stream of profits. We have also shown that this general insight holds not only for the simple price competition game but for oligopoly games more generally. So far, we have restricted the analysis to grim trigger strategies. However, firms may want to use different strategies, raising the question of which are the best strategies in such an infinitely repeated game in order to support full or partial collusion. We turn to this issue in the following subsection.

14.2.2 Optimal punishment of deviating firms

We concluded the previous subsection by comparing the sustainability of tacit collusion under price and quantity competition. One important difference between the two settings arises from the strength of punishment. Under price competition, reversion to the Nash equilibrium generates zero profits for all firms, which provides the most severe credible punishment.[17] In contrast, under quantity competition, reversion to the Nash equilibrium gives rise to positive profits. It thus seems possible to design more severe punishment schemes enabling the firms to sustain collusive outcomes over a larger set of discount factors. On the contrary, the punishment must be credible, in the sense that firms must find it optimal to abide by it. With regard to the grim trigger strategy, the reversion to the Nash equilibrium forever ensures this credibility. With more severe punishments, credibility can be restored by shortening the punishment phase and eventually coming back to the collusive outcome, so as to reward firms that punish a deviator.

To formalize the basic idea, we consider the following *stick-and-carrot strategies*, in which deviators are punished for only one period:

(i) Start the game by playing the collusive output q^*, as prescribed by the collusive agreement.

(ii) Cooperate as long as the collusive output has been observed in all preceding periods.

(iii) If one of the players deviates from the collusive agreement at period t, play $\hat{q}$ at period $t+1$ (punishment phase) and return to the collusive agreement at period $t+2$.

(iv) If one of the players chooses a quantity $q \neq \hat{q}$ during the punishment phase, start the punishment phase again at the following period.

[17] Under cost asymmetries things are more tricky. See Miklós-Thal (2011).

With this strategy, the punishment has a simple two-phase stick-and-carrot structure: the profit reduction following a deviation is used as a stick, while the promise to return to the collusive outcome after just one period constitutes the carrot. In the following we explain how this strategy is used to sustain collusion in general; we then use the specific example of a linear Cournot duopoly to show how the stick-and-carrot strategy expands the firms' ability to sustain collusive outcomes.

General analysis

It can be shown that in the set of stationary symmetric strategies the stick-and-carrot strategy is optimal when the quantity $\hat{q}$ is chosen to maximize the scope for collusion.[18] For a given discount factor δ, we need to characterize the best collusive sustainable output q^* and the punishment output $\hat{q}$. The punishment output $\hat{q}$ is set to minimize the deviator's profit under the constraint that deviations from the punishment do not pay for any of the firms.

To carry out the general analysis, we need some additional notation. We consider n firms that produce a homogeneous good at constant marginal cost c. The inverse demand is given by $P(q)$. We assume that there is a unique monopoly quantity defined by $q^m = \arg\max_q q(P(q) - c)$, and that $q(P(q) - c)$ is monotonically increasing until q^m and monotonically decreasing after q^m. Monopoly profits are denoted $\pi^m = q^m(P(q^m) - c)$. If $q_i = q^m/n$ for all i, then each firm obtains the collusive profit $\pi^c = \pi^m/n$. We also suppose that the static Cournot oligopoly game has a unique symmetric pure strategy equilibrium, with quantities $q^n \neq q^m/n$ and profit π^n. If all other firms choose a combined quantity $(n-1)q$, let $\pi^d(q)$ denote the profit obtained by a firm that responds optimally to each competitor's quantity q. Formally,

$$\pi^d(q) = \max_z z(P(z + (n-1)q) - c).$$

We also define $\pi(q)$ as the profit obtained by a firm when all firms produce the same quantity q:

$$\pi(q) = q(P(nq) - c).$$

By definition, $\pi(q^m/n) = \pi^m/n$. Notice also that, at the symmetric Cournot equilibrium, $\pi(q^n) = \pi^d(q^n) = \pi^n$.

Using this notation, we express the present discounted value of complying with the punishment as $V^p \equiv \pi(\hat{q}) + [\delta/(1-\delta)]\pi(q^*)$. That is, in the period that follows the deviation, both firms choose the punishment output $\hat{q}$, which gives them profits equal to $\pi(\hat{q})$, and from there on, they return to the collusive agreement by choosing both the collusive output q^*, which gives them the collusive profits $\pi(q^*)$. On the contrary, a deviation from the punishment would yield $\pi^d(\hat{q}) + \delta V^p$. That is, deviation yields immediate profits of $\pi^d(\hat{q})$, which stems from the best response to the rival's output $\hat{q}$, and triggers a new start of the punishment phase, which generates a stream of profits given by V^p. Hence, no deviation occurs as long as $V^p \geq \pi^d(\hat{q}) + \delta V^p$, which is equivalent to

$$\delta\left[\pi(q^*) - \pi(\hat{q})\right] \geq \pi^d(\hat{q}) - \pi(\hat{q}). \quad (14.5)$$

This *credibility condition* is easily interpreted: the cost of not complying with the punishment, which stems from the one-period delay before returning to the collusive outcome, must be larger than the immediate gain of deviating from the punishment. For the punishment to be

[18] See Abreu (1986, 1988).

the harshest possible, condition (14.5) must hold as equality. That is, $\hat{q}$ is found as the solution to $\delta[\pi(q^*) - \pi(\hat{q})] = \pi^d(\hat{q}) - \pi(\hat{q})$.

It must also be checked that it does not pay to deviate from the cooperative phase. We only need to look one period ahead as firms return to the collusive outcome after one period of punishment. Complying with the collusive agreement yields $(1+\delta)\pi(q^*)$, while deviating yields $\pi^d(q^*) + \delta\pi(\hat{q})$, that is, the profits resulting from the best deviation from q^* followed by the profits obtained in the punishment phase. Hence, firms abide by the collusive agreement as long as $(1+\delta)\pi(q^*) \geq \pi^d(q^*) + \delta\pi(\hat{q})$, or equivalently

$$\delta\left[\pi(q^*) - \pi(\hat{q})\right] \geq \pi^d(q^*) - \pi(q^*). \tag{14.6}$$

This *sustainability condition* has a similar interpretation as the previous credibility condition: the cost of deviation (incurred next period) must be larger than the immediate gain of deviation.

Full collusion is sustainable if condition (14.6) is satisfied for $q^* = q^m/2$ and with $\hat{q}$ being the most severe punishment obtained by solving (14.5) with equality. Otherwise, only partial collusion can be sustained, and the best collusive output is found as the solution q^* to $\delta[\pi(q^*) - \pi(\hat{q})] = \pi^d(q^*) - \pi(q^*)$.

Application to the linear Cournot duopoly

We return to the special case that $P(q) = a - q$ and that duopolists have the same constant marginal cost c. We use the definition of $\pi(q)$ and set $n = 2$ in expressions (14.3) and (14.4) to find

$$\pi(q^*) = \tfrac{1}{8}(a-c)^2, \quad \pi(\hat{q}) = (a - c - 2\hat{q})\hat{q}, \quad \text{and} \quad \pi^d(\hat{q}) = \tfrac{1}{4}(a - c - \hat{q})^2.$$

Then, condition (14.5) holding as equality can be rewritten as

$$\delta\left[\tfrac{1}{8}(a-c)^2 - (a-c-2\hat{q})\hat{q}\right] = \tfrac{1}{4}(a-c-\hat{q})^2 - (a-c-2\hat{q})\hat{q}.$$

This quadratic form in $\hat{q}$ admits two roots. As punishment requires expanding output, we select the larger root.[l] This is admissible as long as the market price remains above zero, that is, $p = a - 2\hat{q} \geq 0$, or $\hat{q} \leq a/2$. Making the computations, we find the following:

$$\hat{q}(\delta) = \begin{cases} \frac{2(3-2\delta)+\sqrt{2\delta}}{2(9-8\delta)}(a-c) & \text{for } \delta < \frac{1}{2}\left(\frac{a+2c}{a+c}\right)^2, \\ \frac{1}{2}a & \text{for } \delta \geq \frac{1}{2}\left(\frac{a+2c}{a+c}\right)^2. \end{cases}$$

For which range of discount factors is full collusion sustainable? For $q^* = q^m/2 = (a-c)/4$, we compute that $\pi^d(q^*) = 9(a-c)^2/64$. Hence, condition (14.6) becomes

$$\delta\left[\tfrac{1}{8}(a-c)^2 - (a-c-2\hat{q}(\delta))\hat{q}(\delta)\right] \geq \tfrac{9}{64}(a-c)^2 - \tfrac{1}{8}(a-c)^2 = \tfrac{1}{64}(a-c)^2.$$

Start with $\delta \geq \frac{1}{2}(\frac{a+2c}{a+c})^2$. Then $\hat{q}(\delta) = a/2$, and the condition simplifies to[m]

$$\delta\left[\tfrac{1}{8}(a-c)^2 + \tfrac{1}{2}ac\right] \geq \tfrac{1}{64}(a-c)^2$$
$$\Leftrightarrow \delta(a+c)^2 \geq \tfrac{1}{8}(a-c)^2.$$

[l] The other root never satisfies the sustainability condition.

[m] We assume here that $c < a/2$, so that the best response to $\hat{q} = a/2$ is a positive quantity; otherwise, the best response is zero, which implies that $\pi^d(\hat{q}) = 0$ and that the condition is less stringent.

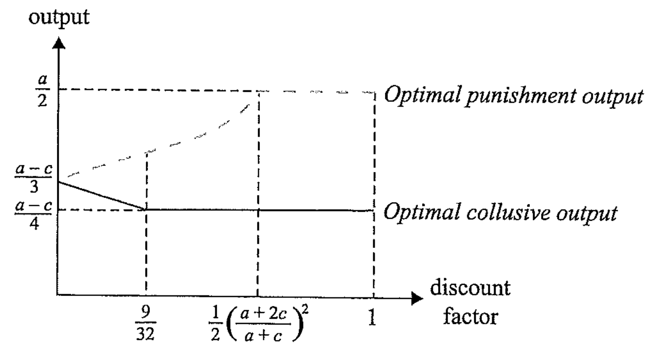

Figure 14.1 Stick-and-carrot strategy in the linear Cournot duopoly

Hence, condition (14.6) becomes $\delta \geq \frac{1}{8}(\frac{a-c}{a+c})^2$, which is implied by condition (14.5) in this case, namely $\delta \geq \frac{1}{2}(\frac{a+2c}{a+c})^2$.

For $\delta < \frac{1}{2}(\frac{a+2c}{a+c})^2$, we ease the computations by setting $x \equiv \sqrt{2\delta}$, so that $\delta = x^2/2$. Using this change of variable and introducing the appropriate value of $\hat{q}\,(\delta)$, we can rewrite condition (14.6) as

$$\frac{x^2}{2}\left[\frac{1}{8}(a-c)^2 - \left(a - c - 2\frac{2(3-x^2)+x}{2(9-4x^2)}(a-c)\right)\frac{2(3-x^2)+x}{2(9-4x^2)}(a-c)\right] \geq \frac{1}{64}(a-c)^2$$

$$\Leftrightarrow \frac{x^2}{2}\left[\frac{1}{8} - \left(1 - 2\frac{2(3-x^2)+x}{2(9-4x^2)}\right)\frac{2(3-x^2)+x}{2(9-4x^2)}\right] \geq \frac{1}{64},$$

which further simplifies to

$$\frac{x^2}{16\,(3-2x)^2} \geq \frac{1}{64} \Leftrightarrow \delta = \frac{x^2}{2} \geq \frac{9}{32}.$$

We thus find that full collusion is sustainable as long as $\delta \geq 9/32 \simeq 0.28$.

What happens for discount factors smaller than 9/32? As indicated above, full collusion (i.e., each firm producing half the monopoly output, $(a-c)/4$) can no longer be sustained. However, partial collusion (i.e., each firm producing more than $(a-c)/4$ but less than the Nash equilibrium output, $(a-c)/3$) can be sustained. To find the best collusive output and the corresponding punishment output, we must solve the system made up of conditions (14.5) and (14.6) holding as equalities. In our linear example, this system admits two solutions. The first solution is, without any surprise, the Nash equilibrium of the static game: $q^* = \hat{q} = (a-c)/3$. Obviously, this solution does not allow firms to sustain collusion. The second solution is the optimal stick-and-carrot strategy:

$$q^* = \frac{1}{27}(9 - 8\delta)(a-c) \quad \text{and} \quad \hat{q} = \frac{1}{27}(9 + 8\delta)(a-c).$$

Notice that the collusive output q^* ranges from the Nash output for $\delta = 0$ to half the monopoly output for $\delta = 9/32$, while the punishment output ranges from the Nash output for $\delta = 0$ to the value $\hat{q}\,(\delta)$ we found above for $\delta = 9/32$. Figure 14.1 represents the optimal stick-and-carrot strategy $(q^*, \hat{q})$ for all discount factors in the linear Cournot duopoly model.

Collecting our previous results, we can conclude the following.

Lesson 14.9 **In the infinitely repeated Cournot quantity-setting game, the set of discount factors that can support full collusion is larger when firms use the optimal punishment of the stick-and-carrot strategy rather than the reversion to the Nash equilibrium of the grim trigger strategy.**

14.2.3 Collusion and multimarket contact

Up to this point we have assumed that firms repeatedly interact in the same market. However, in many real-world cases, firms face largely the same competitors in several markets. This leads to the question of whether multimarket contact facilitates collusion. Identifying the types of environment in which collusion is likely to occur is important for antitrust authorities, as they can only dedicate limited resources to the detection of collusive behaviour.

Multimarket contact has contrasting effects on the sustainability of collusion: on the one hand, it makes deviation from a collusive outcome more profitable (as firms can deviate on all markets at the same time); but on the other hand, it also makes deviation more costly (as deviators would also be punished on all markets). These two opposite effects cancel out when markets are identical, firms are identical and technology exhibits constant returns to scale. In this case, multimarket contact does not facilitate collusive behaviour. However, when one of these conditions is not met, firms may use multimarket contact to pool their incentive constraints across markets and thereby improve their ability to sustain collusive outcomes. Below we analyse different settings that illustrate this point.[19] Case 14.3 gives some evidence that multimarket contact does indeed facilitate collusion.

Case 14.3 Multimarket contact in the US airline industry[20]

The airline industry appears as an ideal candidate for the empirical testing of the effects of multimarket contact on pricing. First of all, firms in the airline industry do indeed compete with each other on several markets. Those markets are easily identified as different city-pair routes. Second, theory predicts (as we show below) that multimarket contact may facilitate collusion (i) when firms differ in their production costs across markets, or (ii) when markets themselves differ. Both conditions are relevant in this industry. Cost differences among firms are likely to result from the hub-and-spoke model, which gives a significant cost advantage to the carrier operating the hub.[n] As for market differences, one observes significant cross-route differences both in the number of operating firms and the rate at which demand is growing. Finally, based on documented evidence, industry experts have claimed that airlines have long

[19] We follow here the seminal analysis of Bernheim and Whinston (1990).

[20] This case is based on Evans and Kessides (1994).

[n] According to this model, an airline uses a central airport (the hub) as a transfer point to get passengers from one city to another (the spokes).

lived by the 'golden rule', whereby airlines refrain from pricing aggressively in a given route for fear of retaliation in another jointly contested route.

To test the effects of multimarket contact, Evans and Kessides (1994) analyse time-series and cross-sectional variability of airline fares in the 1000 largest city-pair routes between 1984 and 1988. Their estimation of a fixed-effects price equation indicates that multimarket contact has a statistically significant and quantitatively important effect on price: fares are, on average, higher on routes where the competing carriers have extensive inter-route contacts. As an illustration, they estimate that moving from the route in their sample with the 25th percentile in contact to a route with the 75th percentile increases prices by 5.1% (which corresponds to an increase of a round-trip ticket price by almost \$13 on the median ticket price in 1988).

Differences between markets

Suppose that each market is characterized by repeated (Bertrand) price competition. As we have seen above, the optimal punishment in this case consists of reverting to the static Bertrand solution forever. Before exploring the effects of multimarket contact, let us look more closely at the factors that influence collusion on each market taken separately. We have already shown above that collusion becomes harder to sustain as the number of firms on the market increases. This is because the gain from deviation increases while the punishment remains the same. A similar intuition underlies the result that a lower frequency of interaction or of price adjustments hinders collusion. Indeed, if firms have to wait longer before interacting again or before adjusting their prices, the punishment following a deviation will come later, thereby allowing the deviator to benefit longer from its cheating behaviour.

To see this, suppose that n firms compete only every k periods (i.e., they interact in periods 1, $k+1$, $2k+1$, etc.). In that case, the present discounted values of abiding by the collusive agreement and of deviating are respectively given by $V_1^C = (\pi^m/n) + \delta^k(\pi^m/n) + \delta^{2k}(\pi^m/n) + \cdots = (\pi^m/n)/(1-\delta^k)$ and $V_1^D = \pi^m$. Suppose alternatively that firms interact in each period but have to fix their prices for k periods (i.e., the price set in period 1 remains valid up to period k; a new price is then set in period $k+1$ and so forth). Here, we have that $V_2^C = (\pi^m/n)/(1-\delta)$ and $V_2^D = \pi^m(1+\delta+\delta^2+\cdots+\delta^{k-1}) + \delta^k \times 0 = \pi^m(1-\delta^k)/(1-\delta)$.

The two conditions for the sustainability of collusion, $V_1^C \geq V_1^D$ and $V_2^C \geq V_2^D$, boil down to the same inequality: $1-\delta^k \leq 1/n$. Therefore, the fully collusive outcome is sustainable if and only if

$$\delta \geq \left(1-\frac{1}{n}\right)^{1/k}.$$

As already argued above, this threshold increases with n. The new result is that the threshold also increases with k. Hence, collusion is easier to sustain in fast-moving markets than in markets where transactions are irregular.

Lesson 14.10 **Firms find it harder to sustain collusion when they interact less frequently or when price adjustments are less frequent.**

To see the effects of multimarket contact, suppose now that two firms compete not only in one but in two markets. Suppose also that the two markets differ along some dimension. In our first illustration, firms interact more frequently on one market than on the other; in our second, one market is more competitive than the other.

Different frequency of interaction. Suppose that firms can change prices more frequently in market 1 than in market 2. To give a concrete specification, suppose that firms can change prices in every period in market 1 ($k = 1$) and only in even periods in market 2 ($k = 2$). The implicit discount factor between periods is δ. Considering markets separately and applying the above analysis, we find that the threshold on the discount factor is $1/2$ in market 1 and $\sqrt{1/2} \approx 0.707$ in market 2. Hence, if $0.5 \leq \delta < 0.707$, collusion is sustainable in market 1 but not in market 2, when considering the two markets separately. This would indeed be the result if we had two different sets of firms in the two markets.

Multimarket contact opens up the possibility that a deviation in market 2 can be (immediately) punished in market 1. With full collusion, discounted profits are $\pi^m/(1-\delta)$. A firm that undercuts makes monopoly profits in market 1 for one period and in market 2 for two periods. Thus the discounted deviation profit is $\pi^m + (1+\delta)\pi^m = (2+\delta)\pi^m$. Such a deviation is not profitable if $2 + \delta \leq 1/(1-\delta)$, which is equivalent to $\delta^2 + \delta - 1 \geq 0$ or $\delta \geq (\sqrt{5}-1)/2 \approx 0.618$. This illustrates how the pooling of incentive constraints across the two markets facilitates collusion.

Different number of firms. We can repeat the analysis by supposing instead that the two markets differ in their number of firms (but are identical otherwise). Suppose that firms A and B are both present in markets 1 and 2, and that firm C is present on market 2 only. Supposing that interaction takes place in every period and considering the two markets separately, we have that for $1/2 \leq \delta < 2/3$, collusion can be sustained in market 1 (where there are two firms) but not in market 2 (where there are three firms and where the threshold on the discount factor is thus $1 - (1/3) = 2/3$). However, firms A and B can take advantage of their multimarket contact to sustain collusion on both markets. The idea is to induce firm C to collude by leaving it a larger market share on market 2, while using the interaction on market 1 as a disciplining device. More precisely, let s denote firm C's share of market 2. Firm C does not deviate as long as $s\pi^m/(1-\delta) \geq \pi^m$, or $s \geq 1 - \delta$. So, firms A and B will leave a share of $1 - \delta$ to firm C and keep a share $\delta/2$ for each of them. Considering the profits to be made on the two markets, A and B prefer not to deviate as long as

$$\frac{1}{1-\delta}\left[\frac{\pi^m}{2} + \frac{\delta\pi^m}{2}\right] \geq 2\pi^m \Leftrightarrow \delta \geq \frac{3}{5}.$$

This new threshold lies between $1/2$ and $2/3$. Hence, for $\delta \in [3/5, 2/3]$, firms A and B are able to sustain collusion in both markets although they would not be able to collude on market 2 were they only active on that market. The reason is that a discount factor strictly larger than $1/2$ gives some slack enforcement power in market 1, which the firms can then use to discipline collusion in market 2.

So, whatever the nature of the differences between markets, the message is the same.

Lesson 14.11 **With multimarket contact on different markets, collusion may become sustainable in several markets, even though deviations would be profitable if firms were active only in one of the markets.**

Differing firms: market-sharing agreements

Reciprocal market-sharing agreements, whereby firms refrain from entering each other's territory, are commonly observed in many industries, as illustrated in Case 14.4 below. Such agreements are likely to occur in situations where firms have separate home markets and incur a transportation cost to serve the other market, or where firms face a fixed cost of production. We develop an example corresponding to the first situation: markets are identical but firms differ in the sense that the 'home' firm has a lower marginal cost than the 'foreign' firm. We show that in a Bertrand model with homogeneous products, multimarket contact may again facilitate collusion in the presence of such cost differences. In particular, collusion is easier to sustain under market-sharing agreements than when firms are present on both markets and assign production quotas.[21]

Case 14.4 Market-sharing agreements in Europe and the USA[22]

Market-sharing agreements have long been held under suspicion by antitrust authorities. In one of the earliest cases litigated under the Sherman Act, the Addyston Pipes Case of 1899, the Supreme Court struck down a group of iron pipe producers which rigged prices on certain markets, and reserved some cities as exclusive domains of one of the sellers.[23] In recent years, the globalization of markets and the deregulation of industries that used to be regulated on a territorial basis (airlines, local telecommunication services and utilities) have increased the scope for explicit or implicit market-sharing agreements.

The European Commission has been particularly aware of the potential risk of market-sharing, as firms that used to enjoy monopoly power in some territories seem reluctant to compete on the global European market. In a landmark case against Solvay and ICI in 1990, the European Commission has established that the two companies had operated a market-sharing agreement for many years by confining their soda-ash activities to their traditional home markets, namely continental Western Europe for Solvay and the UK for ICI. It was also found that over many years, all the soda-ash producers in Europe accepted and acted upon the 'home market' principle, under which each producer limited its sales to the country or countries in which it had established production facilities.[24] In the USA, the Telecommunications Act of 1996 was designed specifically to encourage regional operators to enter each other's markets. It appears today that both industries are still dominated by a handful of dominant companies, each with highly clustered regional monopolies.

Antitrust authorities reacted by issuing new guidelines that emphasize market-sharing agreements as an alternative form of collusion. For example, in its 1999 merger guidelines, the Irish Competition Authority states that: 'As an alternative to a price-fixing cartel, firms ... may divide up the country between them and agree not to sell in each other's designated area. ... At its simplest, a market-sharing cartel may be no more than an agreement among firms not to approach each other's customers or not to sell to those in a particular area. This

[21] Belleflamme and Bloch (2008) show that this result is only partly true when firms compete in quantities.
[22] This case draws from Belleflamme and Bloch (2004).
[23] See Scherer and Ross (1990), pp. 318–319.
[24] See Official Journal L 152, 15/06/1991, pp. 1–15.

may involve secretly allocating specific territories to one another or agreeing on lists of which customers are to be allocated to which firm.' (Irish Competition Authority, 1999)

We consider again two markets (1 and 2) and two firms (A and B). Suppose that firm A is installed in market 1, while firm B is installed in market 2. Each firm produces a homogeneous good at constant marginal cost c and faces a transportation cost of τ to move one unit of output from its 'home' market to the 'foreign' market. Suppose also that demand on both markets is given by $Q(p) = a - p$. The monopoly price for the home firm is equal to $p^m(c) = (a+c)/2$. To ensure that competition between the two firms is viable, we assume that the marginal cost of the foreign firm is smaller than the monopoly price: $c + \tau < (a+c)/2$, or $2\tau < a - c$.

As a benchmark, we first examine the optimal collusive outcome that would prevail if firms were only competing in one market. As we assume Bertrand competition, the optimal punishment consists for both firms in setting their price equal to c in every future period following a deviation; the home firm makes all the sales and both firms earn zero discounted profits. Let s_h and s_f denote the respective market shares of the home and the foreign firm (by definition, $s_h + s_f = 1$). If the collusive price is $p \geq c + \tau$, a deviating firm with cost c_i will slightly undercut and achieve an immediate profit equal to $\pi^d(p, c_i) = (p - c_i)(a - p)$, where $c_i = c$ for the home firm and $c_i = c + \tau$ for the foreign firm. Hence, both firms abide by the collusive agreement as long as

$$\begin{cases} \text{(home firm) } \frac{1}{1-\delta} s_h (p-c)(a-p) \geq (p-c)(a-p) \Leftrightarrow s_h \geq 1-\delta, \\ \text{(foreign firm) } \frac{1}{1-\delta} s_f (p-c-\tau)(a-p) \geq (p-c-\tau)(a-p) \Leftrightarrow s_f \geq 1-\delta. \end{cases}$$

We can sum the two conditions and conclude that only if $\delta \geq 1/2$ can the firms sustain collusive prices above $c + \tau$. It must be noted, however, that a large enough market share must be allocated to the inefficient foreign firm to keep it from deviating; in particular, it must be that $s_f \geq 1 - \delta$.

Consider now the two-market setting. We focus on symmetric collusive outcomes whereby both firms set the same price p on both markets and the home firm receives a share s_h. For a given $p \geq c + \tau$, it is easy to see that the best collusive outcome involves $s_h = 1$, which implies that each firm withdraws completely from the foreign market.° Then, the best collusive price is $p^m(c) = (a+c)/2$. It follows that the present discounted value of abiding by the market-sharing agreement is

$$V^C = \frac{1}{1-\delta} \frac{(a-c)^2}{4}.$$

The best deviation consists of entering the foreign market and undercutting the home firm, thereby obtaining an immediate profit of $(p^m(c) - c - \tau)(a - p^m(c))$; meanwhile, the firm keeps the monopoly profit on its home market. As before, the punishment that follows the deviation yields a continuation profit of zero. So, the present discounted value of deviation is

$$V^D = \frac{(a-c)^2}{4} + \frac{(a-c)(a-c-2\tau)}{4}.$$

° The collusive profit is $\pi^c(s_h) = s_h(p-c)(a-p) + (1-s_h)(p-c-\tau)(a-p) = (a-p)(p-c-\tau+\tau s_h)$, which increases with s_h.

Therefore, a market-sharing agreement specifying that each firm sets $p^m(c)$ on its home market and does not enter the foreign market can be sustained if $V^C \geq V^D$, or

$$\frac{\delta}{1-\delta}\frac{(a-c)^2}{4} \geq \frac{(a-c)(a-c-2\tau)}{4} \Leftrightarrow \delta \geq \frac{1}{2}\frac{a-c-2\tau}{a-c-\tau}.$$

It is clear that the latter threshold is smaller than $1/2$, which shows the following result.

Lesson 14.12 **The optimal market-sharing agreement can be sustained over a larger set of discount factors than the most profitable collusive outcome that firms can achieve when they are present on both markets.**

14.2.4 Tacit collusion and cyclical demand

Many markets are characterized by demand fluctuations. We therefore extend the single-market analysis by considering two demand states. From the point of view of the firm, demand is either good, $Q_G(p)$, or bad, $Q_B(p)$, with $Q_G(p) > Q_B(p)$ for all p. The former situation can be associated with a boom phase of the industry and the latter with a recession phase. To the extent that there is a comovement between the industry under consideration and the economy in general, industry demand expands in an economic boom.

In this subsection, we maintain the assumption that firms observe the state of the demand.[25] They are thus able to detect any deviation from the collusive behaviour by any other firm. For the sake of simplicity, suppose that the good demand state occurs with probability 1/2. The realization of demand is independent over time.[26] In each period, firms learn the state of demand before prices are set.

We want to describe a collusive outcome that can be supported as a (subgame-perfect) equilibrium and is the most preferred outcome by the firms. This means that we look for a pair of prices (p_B, p_G) such that (1) p_B is set in the bad demand state and p_G is set in the good demand state, (2) deviations from (p_B, p_G) are not profitable and (3) the equilibrium is such that there does not exist another equilibrium which is preferred by both firms. The expected present discounted profit along the equilibrium path is

$$\begin{aligned}V^C &= \sum_{t=0}^{\infty}\delta^t\left(\frac{1}{2}\frac{Q_B(p_B)}{2}(p_B-c)+\frac{1}{2}\frac{Q_G(p_G)}{2}(p_G-c)\right)\\ &= \frac{1}{1-\delta}\left(\frac{1}{2}\frac{Q_B(p_B)}{2}(p_B-c)+\frac{1}{2}\frac{Q_G(p_G)}{2}(p_G-c)\right).\end{aligned}$$

To support the preferred equilibrium, firms use trigger strategies so that the punishment for deviation is maximal. Recall that this means, after a deviation from (p_B, p_G), firms set the competitive (Nash equilibrium) price c forever.

[25] The analysis is based on Rotemberg and Saloner (1986).
[26] Allowing for correlated demand would substantially complicate the analysis without affecting our main insights.

Full collusion

We start by providing the conditions needed to sustain the fully collusive outcome in equilibrium. In fact, it is only a simple extension of the result in a world with demand certainty. The fully collusive outcome consists of both firms charging the respective monopoly price p_s^m in state $s = 1, 2$. That is, $p_s^m = \arg\max_{p_s} \pi_s(p_s) = \arg\max_{p_s} (p_s - c)\, Q_s(p_s)$. Let $\pi_s^m \equiv (p_s^m - c)Q_s(p_s^m)$ denote the monopoly profit in state s. Expected present discounted profit for prices (p_B^m, p_G^m) can be written as

$$V^m = \frac{1}{1-\delta}\frac{(\pi_B^m + \pi_G^m)}{4}. \tag{14.7}$$

A deviation in state s leads to (almost) the monopoly profit in that period and to zero profits in all future periods. Thus, like in the analysis with demand certainty, a deviation leads to a short-term gain and a long-term loss. The upper bound for expected present discounted profit of a deviation is thus $V_s^D = \pi_s^m$ in state s. Along the equilibrium path, the firm would make profit $\pi_s^m/2$ in state s.

For a deviation to be not profitable, one must have

$$V_s^D \leq \frac{\pi_s^m}{2} + \delta V^m \Leftrightarrow \frac{\pi_s^m}{2} \leq \delta V^m. \tag{14.8}$$

Using (14.7), we can rewrite (14.8) as

$$\frac{\pi_s^m}{2} \leq \frac{\delta}{1-\delta}\frac{\pi_B^m + \pi_G^m}{4} \Leftrightarrow 2(1-\delta)\,\pi_s^m \leq \delta\pi_B^m + \delta\pi_G^m.$$

Note that since $\pi_B^m < \pi_G^m$, condition (14.8) is more stringent in the good demand state as deviation is more profitable while the severity of the punishment is independent of the demand state. Collusion is thus sustainable as long as $\pi_G^m/2 \leq \delta V^m$. Here, the punishment entails the loss of an average of high and low profits, so that the punishment is less severe than if the good demand state persisted, as would be the case with demand certainty. In other words, *the fully collusive outcome is more difficult to sustain than under demand certainty*. In the good demand state, the inequality simplifies to $2\pi_G^m \leq \delta\pi_B^m + 3\delta\pi_G^m$. Solving for δ, we obtain a critical value δ^0 above which the fully collusive outcome can be supported,

$$\delta \geq \delta^0 \equiv \frac{2\pi_G^m}{\pi_B^m + 3\pi_G^m} = \frac{1}{1 + \underbrace{\frac{\pi_B^m + \pi_G^m}{2\pi_G^m}}_{\in(\frac{1}{2},1)}}.$$

This critical value lies within the interval (1/2, 2/3).

Lesson 14.13 **Under demand uncertainty, the critical discount factor above which the fully collusive outcome can be sustained is larger than under demand certainty.**

As we have seen, a firm is more tempted to undercut when the demand state is good. (Recall that condition (14.8) holds if and only if it holds in the good demand state.) This suggests that for lower discount factors, full collusion cannot be sustained in the good demand state, whereas it can be sustained in the bad demand state. We now turn to this type of partial collusion.

Partial collusion

We now provide conditions such that a partially collusive outcome can be sustained in equilibrium. To this end, we consider discount factors $\delta \in [\frac{1}{2}, \delta^0)$. We choose (p_1, p_2) so as to maximize expected profits subject to the undercutting constraints (i.e., incentive constraints). That is, we solve the problem

$$\max_{p_B, p_G} \frac{1}{1-\delta}\left(\frac{1}{2}\frac{\pi_B(p_B)}{2} + \frac{1}{2}\frac{\pi_G(p_G)}{2}\right)$$

subject to

$$\frac{\pi_B(p_B)}{2} \leq \frac{\delta}{1-\delta}\left(\frac{1}{2}\frac{\pi_B(p_B)}{2} + \frac{1}{2}\frac{\pi_G(p_G)}{2}\right), \tag{14.9}$$

$$\text{and } \frac{\pi_G(p_G)}{2} \leq \frac{\delta}{1-\delta}\left(\frac{1}{2}\frac{\pi_B(p_B)}{2} + \frac{1}{2}\frac{\pi_G(p_G)}{2}\right). \tag{14.10}$$

It is still true that the temptation to undercut is higher when the demand state is good. This implies that the binding constraint is (14.10). Constraint (14.9) can thus be ignored. Taking this into account and rearranging terms, we can rewrite the problem as

$$\max_{p_B, p_G} \pi_B(p_B) + \pi_G(p_G) \quad \text{subject to} \quad \pi_G(p_G) \leq K\pi_B(p_B),$$

where $K \equiv \delta/(2-3\delta)$ with $K \geq 1$ for $\delta \geq 1/2$. Note that the constraint is relaxed if π_B is increased. Since maximal profit in the bad demand state is $\pi_B(p_B^m) = \pi_B^m$, the high-demand price $\widetilde{p}_G$ is then chosen so as to satisfy $\pi_G(p_G) = K\pi_B^m$. This implies that for $\delta \in [\frac{1}{2}, \delta^0)$, firms partially collude with $(p_B^m, \widetilde{p}_G < p_G^m)$: in the bad demand state, firms charge the monopoly price; in the good demand state, firms set prices more aggressively than under monopoly.

In particular, the result is compatible with $p_G < p_B^m$, in which case firms price lower in the good than in the bad demand state (but this is only a possibility not an implication of the model). In other words, the model can generate lower prices after a positive demand shock and countercyclical prices and markups.[27]

Lesson 14.14 **When full collusion cannot be sustained, firms may partially collude by setting the respective monopoly price in the bad demand state and setting a price lower than the respective monopoly price in the good demand state. This result is compatible with countercyclical prices and markups.**

---

[27] Note that the Rotemberg–Saloner model is not the only explanation of countercyclical markups. Another explanation is that demand may be less elastic during recessions, leading to higher markups in monopolistic or imperfectly competitive markets in the absence of collusion; see, for example, Stiglitz (1984) and Bils (1989). Also, capital market imperfections can generate countercyclical markups; see, for example, Chevalier and Scharfstein (1996). According to this explanation, firms have low cash flows and greater difficulties in obtaining external funding during recessions. As a response, firms try to boost current profits by increasing price and care less about building market share.

A number of empirical observations can be related to the model. In the market for a particular drug, namely antibiotic tetracycline, pure discipline among firms broke down when the Armed Service Medical Procurement Agency placed a large order in October 1956, which can be interpreted as a positive demand shock.[28] Also, it has been observed that the price of cement tends to move countercyclically.[29] The suspicion is that a cartel cannot sustain full collusion in a good demand state.

14.2.5 Tacit collusion with unobservable actions

We now present a model (of price competition) in which firms do not directly observe deviations from the collusive outcome.[30] Firms cannot make direct inferences about deviation from market outcomes because we postulate that there is uncertainty in the market which is not generated by actions taken by the firms. In particular, our key assumption is that firms do not observe their rivals' prices but infer them (imperfectly) from their own demand. Furthermore, firms do not observe the state of demand before each period.

For simplicity, we consider collusion among two firms. As before, products are perfect substitutes produced at c and there are two demand states, a good and a bad one. In the bad demand state, demand is assumed to be zero, whereas in the good demand state, there is positive demand at $p = c$, that is, $Q(c) > 0$, and demand is strictly decreasing (at prices with strictly positive demand, $Q(p) > 0$).

The high-demand state occurs with probability $(1 - \alpha)$. To avoid complications, we assume that realization of demand is i.i.d. over time. Denote the monopoly price in the good demand state by p^m and the corresponding monopoly profit by π^m. If a firm cannot sell at a particular point in time, it does not know whether this is due to the fact that demand is zero or because the rival prices lower so that it does not obtain any demand. In such a situation, the firm facing zero demand has to solve a nontrivial signal extraction problem: if demand is zero, there is no reason to suspect that the competitor deviated from the collusive outcome, whereas if the demand state is good and the competitor deviated from the collusive price, such behaviour should be punished.[p]

Consider the infinitely repeated Bertrand game under demand uncertainty, in which future profits are discounted by δ. We propose the following intertemporal strategies for both firms:

1. Start with the collusive phase and charge price p^m until one firm makes zero profit (which along the equilibrium path, occurs in the low demand state).
2. If one firm makes zero profit, this triggers a punishment phase for $T \in \mathbb{N}_+ \cup \{\infty\}$ periods in which each firm charges c. After T periods, firms return to step 1.

Clearly such strategies lead to higher profits than atemporal profits under Bertrand competition, which would be zero. We have to show that such strategies constitute equilibrium

[28] See Scherer (1980).
[29] See Rotemberg and Saloner (1986).
[30] The analysis is based on Green and Porter (1984).
[p] Note the following: in a situation in which at least one firm faces zero demand, it is common knowledge that at least one firm makes zero profit.

strategies. We construct T such that the expected present discounted value is maximized under some constraints that guarantee that the strategies constitute an equilibrium.

In the punishment phase, a deviation is not profitable for either of the two firms since its profits cannot become strictly positive. Consider the collusive phase. Let V^C denote the expected present value of a firm's profit from date t presuming that the game is in the collusive phase. With probability $(1-\alpha)$, the demand state is good so that firms share monopoly profits and continue in the collusive phase. With the remaining probability, profits are zero and firms enter the punishment phase, where V^P denotes the corresponding expected present value.[q] Hence,

$$V^C = (1-\alpha)\left(\tfrac{1}{2}\pi^m + \delta V^C\right) + \alpha(0 + \delta V^P). \tag{14.11}$$

The present value when entering the punishment phase is determined by the sum of zeros for T periods, which is the length of the punishment phase, and the present value in the collusive phase discounted by δ^T because after T periods firms return to the collusive phase; that is,

$$V^P = \underbrace{0 + 0 + \cdots + 0}_{T\text{ periods}} + \delta^T V^C. \tag{14.12}$$

Firms must not have an incentive to deviate in the collusive phase. Therefore, the collusive profit V^C must be higher than the profit it earns if it undercuts the other firm slightly. In the latter case a firm obtains the full monopoly profit but then finds itself in the punishment period for T periods. Hence, the incentive constraint reads

$$V^C \geq (1-\alpha)(\pi^m + \delta V^P) + \alpha\delta V^P.$$

Using equation (14.11), one can rewrite the condition as

$$(1-\alpha)\,\delta(V^C - V^P) \geq (1-\alpha)\,\tfrac{1}{2}\pi^m, \tag{14.13}$$

where the left-hand side is the expected long-term loss from the deviation, and the right-hand side is the expected short-term gain from the deviation. Hence, to avoid undercutting, V^P must be sufficiently lower than V^C. This means that the punishment must be sustained for sufficiently many periods. To obtain explicit conditions for our parameters, we need to solve the system of equations given by (14.11) and (14.12). Substituting the value of V^P into (14.11), we obtain the value in the collusive phase

$$\begin{aligned} V^C &= (1-\alpha)\left(\tfrac{1}{2}\pi^m + \delta V^C\right) + \alpha\delta^{T+1} V^C \\ &\Leftrightarrow \left[1 - (1-\alpha)\,\delta - \alpha\delta^{T+1}\right] V^C = (1-\alpha)\,\tfrac{1}{2}\pi^m \\ &\Leftrightarrow V^C = \frac{(1-\alpha)\,\frac{1}{2}\pi^m}{1-(1-\alpha)\,\delta - \alpha\delta^{T+1}}. \end{aligned} \tag{14.14}$$

Plugging this value into (14.12) gives the value when entering the punishment phase,

$$V^P = \delta^T \frac{(1-\alpha)\,\frac{1}{2}\pi^m}{1-(1-\alpha)\,\delta - \alpha\delta^{T+1}}. \tag{14.15}$$

[q] Since we are analysing a stationary game, V^P and V^C do not need a time index.

We can now substitute expressions (14.14) and (14.15) into the incentive constraint (14.13):

$$(1-\alpha)\,\delta(1-\delta^T)\frac{(1-\alpha)\,\frac{1}{2}\pi^m}{1-(1-\alpha)\,\delta-\alpha\delta^{T+1}} \geq (1-\alpha)\,\tfrac{1}{2}\pi^m$$

$$\Leftrightarrow (1-\alpha)\,\delta-(1-\alpha)\,\delta^{T+1} \geq 1-(1-\alpha)\,\delta-\alpha\delta^{T+1}$$

$$\Leftrightarrow 2\,(1-\alpha)\,\delta+(2\alpha-1)\,\delta^{T+1} \geq 1. \tag{14.16}$$

The highest profits for the firms are thus the maximum of the expected present value subject to the constraint that no firm has an incentive to undercut in the collusive phase: max V^C subject to (14.16). Note that V^C decreases in T, as can be seen from (14.14). Therefore the solution to this problem is the lowest possible T that satisfies the incentive constraint (14.16). Clearly, not to punish (i.e., $T = 0$) always violates the incentive constraint. This means that a positive length of punishment phase is needed.

Whether collusion can be sustained with punishment depends also on the probability of the bad demand state. In particular, if the bad demand state is sufficiently likely, collusion breaks down. To see this, suppose that $\alpha \geq 1/2$. Consider constraint (14.16), which we can rewrite as

$$2\,(1-\alpha)\,\delta(1-\delta^T)+\delta\delta^T \geq 1.$$

For $\alpha \geq 1/2$, the left-hand side of the inequality is weakly less than $\delta(1-\delta^T)+\delta\delta^T=\delta<1$, which is a contradiction. Hence, the above proposed strategies do not constitute an equilibrium for $\alpha \geq 1/2$ and collusion cannot be sustained.

The left-hand side of the inequality increases with T only if $\alpha < 1/2$ and is maximal for $T=\infty$. To make at least maximal punishment ($T=\infty$) sustainable, one must have $2\,(1-\alpha)\,\delta \geq 1$, which is equivalent to

$$(1-\alpha)\,\delta \geq 1/2. \tag{14.17}$$

That is, for a given discount factor δ, the good demand state must be sufficiently likely (i.e., $1-\alpha$ must be large enough). Similarly, for a given $\alpha < 1/2$, the discount factor must be sufficiently close to 1. If $(1-\alpha)\,\delta > 1/2$, the optimal length of punishment is given by the lowest T that satisfies (14.16). In sum, *if* $(1-\alpha)\,\delta > 1/2$, *there is an equilibrium with the property that phases of collusion and phases of punishment alternate, where each punishment phase is initiated by a bad realization of demand within a collusion phase.* We summarize our analysis with the following lesson.

Lesson 14.15 **Even if firms cannot observe deviations of other firms from equilibrium play, collusion can still be supported to some extent. However, the conditions for collusion to be sustainable are stricter than in a world in which deviations can be immediately observed and punished. In addition, profits are lower.**

With perfect observability, if firms can quickly punish deviators, a deviation becomes less attractive, as the short-term benefit is reduced and the long-term loss increased. Thus, flexibility makes collusion easier to sustain. By contrast, under imperfect monitoring, flexibility may harm collusion. Intuitively, if firms are quite inflexible, by default the colluding parties collect additional pieces of information prior to any actions. Therefore, they receive more precise signals. Not being able to punish quickly may then facilitate collusion.[31]

14.3 Detecting and fighting collusion

In the previous two sections, we have tried to understand under which conditions collusive conduct (be it tacit or explicit) is viable and sustainable. This does not tell us, however, how to use this knowledge to design policy enforcement. A priori, detecting and fighting collusion seem to be fraught with difficulty. We have indeed seen in the previous section that collusive outcomes can be sustained even without explicit agreements. We may therefore ask why competition authorities mainly try to uncover explicit agreements such as price-fixing cartels if firms have other means to sustain collusion.

The answer to the latter question is that tacit collusion without explicit agreements may be hard to sustain because firms may lack information to detect deviations from the collusive outcome and because firms have to coordinate their actions (to select a collusive outcome among the multiple non-cooperative equilibria, and to use an appropriate punishment strategy). Compared with tacit collusions, cartels allow for better monitoring of other firms. In particular, managers may exchange price and output levels between firms so that deviations from collusion can more easily be detected, as illustrated in Case 14.5.

Case 14.5 The vitamin cartels (3)

The different vitamins cartels (see the description in Case 14.1) all operated similarly. Regular meetings allowed the cartel members to monitor and enforce agreements. In particular, information on sales volumes and prices of the companies were exchanged monthly or quarterly. In addition, 'a formal structure and hierarchy at different levels of management, often with overlapping membership at the most senior levels' was established.[32]

Another argument for explicit collusion is that in the presence of demand uncertainty, information sharing within a cartel may be required for firms to be able to observe price or output decisions of other firms. As we have seen in Subsection 14.2.5, if firms face stochastic demand and do not observe the other firms' actions directly, they often have to tackle a signal extraction problem and may misinterpret the action of other firms. To avoid the complete breakdown of collusion, firms must punish actions that, in the absence of the information problem, would not have to be

31 See Abreu, Milgrom and Pearce (1991) and, in particular, Sannikov and Skrzypacz (2007).

32 See European Commission (2001b, p. 2). The members of the cartel were thus able to allocate sales quotas, effectively cementing their market shares, and to agree on target prices including simultaneous price increases. The actions of the different cartels were also coordinated between each other and sometimes even the same management personnel of the companies was involved. Hoffmann-La Roche, as the largest vitamin producer, played a leading role, for example in the negotiations held in Japan and the Far East it represented all European manufacturers. Together with BASF it was alleged to also have convinced the Japanese vitamin producer Eisai to participate in the cartels.

punished. As a consequence, firms make lower profits and the conditions for some type of collusion to be feasible at all are more demanding than in the absence of the signal extraction problem. Thus, monitoring and information-sharing can help to sustain (full) collusion.[33] In addition, in a formal cartel, the coordination on particular actions is easier to achieve. This tends to increase equilibrium profits. Unfortunately, there is no good theory around that allows us to address the advantages of formal agreements over implicit understanding with respect to the coordination problem.[34] Nevertheless, the analysis of the two previous sections leads to the following lesson.

Lesson 14.16 **Without the information-sharing within a cartel, collusion is more likely to be infeasible. Even if it is feasible, collusion may not be supported over the whole time horizon but firms may alternate between collusive and punishment phases (of varying length) in which they switch between a high and a low price.**

The previous discussion suggests that communication is central to collusion. Because of the importance of communication between firms, collusion might leave significant pieces of evidence: permanent records of meetings or agreements (on paper, fax, emails or computer hard-drives) may have been kept, telephone conversations may have been tapped. In the absence of such evidence, detecting and legally proving the existence of a cartel is extremely difficult, as we explain in Subsection 14.3.1. Competition authorities have thus designed policies so as to encourage cartel members to bring evidence to the authorities by themselves. We examine these so-called 'leniency programmes' in Subsection 14.3.2.

14.3.1 The difficulty in detecting collusion

In Chapter 3, we have described several ways to measure price–cost margins in a given industry. One might be tempted to use high price–cost margins as evidence of collusion. Yet, this approach is mistaken: a high price–cost margin simply indicates market power, and we have seen in Part III that there are various sources of market power absent collusion (e.g., product differentiation, advertising, search and switching costs). Therefore, it is not true that a high price–cost margin proves collusion. By the way, the reverse statement is not true either: collusion does not necessarily imply a 'high' price–cost margin. To see this, think of symmetric firms competing à la Bertrand: absent collusion, the price–cost margin would be equal to zero; with collusion, firms might achieve some positive margin, but this margin could be equivalent to that prevailing in a non-colluding industry with limited capacity. In that case, an observer who wrongly thinks that the relevant benchmark is one with limited capacity would infer from the observed price–cost margin that there is no collusion.

[33] Cartel members may be able to effectively communicate private information to each other, for example, about cost levels. If firms talk to each other, they can agree how to split the market. This may allow them to increase productive efficiency because the cartel tends to assign a larger market share to members who happen to have lower costs. For a dynamic analysis, see Athey and Bagwell (2001).

[34] The game-theoretic analysis of cheap talk is closest to this issue. For a discussion of its applicability to the analysis of cartels, see Whinston (2006).

Hence, to detect collusion, we should not only look at the level of the price–cost margin. What should raise suspicions about collusion is a sharp increase in the price–cost margin, because, if market conditions do not change, such an increase is difficult to rationalize in the absence of a cartel. The question then is: How can collusion be detected? Four methods have been described in the literature.[35] A first method asks whether the observed firms' behaviour is consistent or not with properties or behaviour that are supposed to hold under a wide class of competitive models. A second method tests for structural breaks in the behaviour of firms, based on the idea that a discrete change in firms' pricing functions might be due to the formation of a cartel (or to its demise). These two methods, however, do not provide any evidence of collusion; they just show that observed behaviour does not seem to correspond to what can be expected from competition. They are thus used primarily as screening methods.

In contrast, the third and fourth methods can serve verification purposes. The third one asks whether the behaviour of suspected colluding firms differs from that of competitive firms. If only a subset of firms in the industry colludes and if these firms can be identified, the comparison between colluding and competitive firms can be carried out directly. Otherwise, one can resort to comparison across markets (when firms collude in some markets but not in others) or across periods (before and after the suspected date of formation of the cartel). Finally, the fourth method examines which model better fits the data, a collusive or a competitive model.

Arguably, these four methods suffer from two general problems. First, the necessary data to identify firms' behaviour is often not available: cost is often unobservable and, in many cases, price and quantity data may not be publicly available. In consequence, competition authorities have to get the relevant data from the firms suspected of collusion. This places them at an informational disadvantage since firms have a clear incentive to misreport their private information so as to disguise a collusive behaviour as a competitive one. The authorities are thereby likely to suffer from what has been coined the *indistinguishability theorem*.[36]

Here is a simple illustration of this problem. Suppose that the industry consists of n symmetric firms: they produce a homogeneous good at the same constant marginal cost c and face the linear inverse demand $P(q) = a - q$. If firms compete à la Cournot, we are familiar by now with the following result: the Nash equilibrium of this linear Cournot model is such that each firm produces $q^n = (a-c)/(n+1)$, resulting in a price of $p^n(c) = (a+nc)/(n+1)$. If all firms collude, they will jointly produce the monopoly output $q^m = (a-c)/2$ and the market price will be $p^m(c) = (a+c)/2$. Suppose now that the competition authorities can estimate the demand intercept a, but cannot exactly observe the firms' marginal cost. All the authorities know is that the true cost c lies somewhere in the interval $[c^-, c^+]$, with $c^+ < a$. Indistinguishability follows if the lowest possible collusive price, $p^m(c^-)$, lies below the highest possible competitive price, $p^n(c^+)$; this is equivalent to[r]

$$\frac{a+c^-}{2} < \frac{a+nc^+}{n+1} \Leftrightarrow c^+ - c^- > \frac{n-1}{2n}(a-c^-).$$

When the latter inequality is satisfied, firms are able to misreport their cost so as to hide their collusive behaviour without arousing suspicions from the authorities. To put some specific numbers to the example, suppose that $n = 3$, $a = 10$, $c^- = 3$ and $c^+ = 7$. One can check that the above

[35] We follow here Harrington (2008).
[36] See Harstad and Phlips (1994).
[r] It is easily checked that this inequality is compatible with $c^+ < a$.

inequality is satisfied with these numbers. Now assume that the true cost is $c = 4$. Then, if firms collude, the observed price is equal to $p^m(4) = 14/2 = 7$. Since the authorities do not know this true value, firms can simply lie and artificially inflate their cost so as to make the observed price compatible with competitive conduct. In particular, by reporting a cost of 6 (which is credible as $c^- < 6 < c^+$), firms would induce the authorities to believe that the observed price results from Cournot competition (as $p^n(6) = 28/4 = 7$) and not from collusion (as $p^m(6) = 8$).

Even if the data about cost, price and quantity are available, all detection methods have to face the additional problem that the estimation of firms' behaviour may be extremely sensitive to the specification of the models. Case 14.6 illustrates the seriousness of this problem.

Case 14.6 The Joint Executive Committee[37]

The *Joint Executive Committee* was a cartel that railroads created in the late nineteenth century to coordinate the rate charged for transporting grain in the USA from Chicago to the East Coast. As the cartel pre-existed the first antitrust regulations issued in the USA (the Sherman Act), firms had no reason to conceal their coordinated activities and the cartel is thus very well documented. The agreement involved allocation of market shares but prices were set individually by the railroads. This opened up the possibility of secret price cuts which, combined with demand fluctuations, endangered the stability of the cartel. This is in line with the predictions of the theory. We showed indeed in Subsection 14.2.5 that there is an equilibrium with the property that phases of collusion and phases of punishment alternate, where each punishment phase is initiated by a bad realization of demand within a collusion phase.

Porter (1983) and Ellison (1994) have tested this hypothesis and tried to assess how successful the cartel was in raising prices during collusion phases. The two authors came to rather different conclusions: according to Porter's estimate, railroads reached price–cost margins consistent with Cournot behaviour; in contrast, Ellison obtains an estimate close to full collusion. This difference in the results may be attributed to differences in the chosen specifications: Ellison extended Porter's analysis by allowing for autocorrelated demand (a high demand today is more likely to generate a high demand tomorrow) and by assuming that the probability of being in a collusion phase tomorrow depends on whether firms are currently in a phase of collusion or of punishment.

We can sum up the previous discussion as follows.

Lesson 14.17 **In the absence of hard evidence of communication among colluding firms, collusion is hard to detect because data on cost, price and quantity is often unavailable or is manipulated. Moreover, even if data is available, the estimation of the firms' behaviour may be extremely sensitive to the specification of the model.**

[37] This case draws from Kühn (2001) and Harrington (2008).

14.3.2 Leniency and whistleblowing programmes

An important aspect for antitrust authorities and courts that investigate alleged cartels is to find evidence for the existence of a cartel that sustains prices above the industry's competitive level. Following the US example, the EU and other developed countries have introduced *corporate leniency* programmes, by which they provide reduced sentences to firms that cooperate with the antitrust authority or the courts and provide evidence on the existence of a cartel and its inner working. These corporate leniency programmes are also extended to individual informants, or *whistleblowers*, by shielding them from criminal sanctions (including jail). These incentives contribute to accomplish the main objective of law enforcement by making cartels less stable and thus reducing the likelihood that a cartel is formed in the first place. An additional objective is the break-up of existing cartels. Here, the social gain from breaking up the cartel has to be compared with the prosecution cost.

The vitamin cartels provide again an illustration of the implementation of leniency programmes.

Case 14.7 The vitamin cartels (4)

For violation of article 81 of the European Community Treaty and Article 53 of the European Economic Area Agreement, eight companies were fined while the remaining five companies evaded their fines as the respective cartels had become time-barred in 2001. The largest fines were awarded to Hoffmann-La Roche with EUR 462 million, BASF with EUR 296.16 million and Takeda Chemical Industries with EUR 37.05 million. As reflected in the size of their fines, Hoffmann-La Roche and BASF were considered the joint leaders and instigators of the cartels with their strong market shares empowering them to implement the anticompetitive agreements. Aventis, however, was the first company to cooperate both with the US Department of Justice and the European Commission and thus received significantly lower fines. Other cartel members that cooperated with those authorities during the investigations also obtained reduced fines according to their level of cooperation. Yet, despite the vitamin cartels being detected and punished both in the USA and Europe, the incentive to form new cartel arrangements apparently remained and led to another fine in 2004 for animal feed vitamins, where EUR 66.34 million were imposed on Akzo Nobel, BASF and UCB (see European Commission, 2004).

While many cartels involve bulk commodities, some cartel cases concern branded consumer goods, as the following case illustrates. This case exemplifies that cooperation by one of the cartel members was key for uncovering an illegal cartel and fining its members.

Case 14.8 The beer cartel in the Netherlands[38]

According to the findings of the European Commission, four Dutch beer brewers formed a cartel on the beer market in the Netherlands, a branded consumer good market, which is

[38] This case is based on European Commission (2007).

normally not seen as a prime candidate for collusion. The cartel coordinated prices and price changes (at the wholesale level). It operated in two market segments: the segment in which consumption is on the premises (bars, restaurants, hotels) and the retail segment, which are mostly supermarkets in the Netherlands. In the former segment brewers coordinated rebate policies. Occasionally, brewers also allocated customers in both market segments, which amounts to market-sharing agreements.

In 2007 the European Commission fined the three leading Dutch brewers Heineken, Grolsch and Bavaria a total of around EUR 274 million for operating this cartel between at least 1996 and 1999. Essential information came from InBev, which was also participating in the cartel (InBev provided information on similar cartels in other European countries). The information provided by InBev led to surprise inspections. The Commission obtained evidence in the form of handwritten notes taken at the unofficial meetings of the cartel members and the proof of dates of secret meetings. Under the Commission's leniency programme, InBev did not have to pay fines.

The model

To evaluate the effects of leniency and whistleblowing programmes, we extend the tacit collusion setting developed in Section 14.2.[39] As before, we define the following per-period (gross) profit levels: when both firms collude, they both obtain π^c; when one firm colludes and the other optimally deviates, the deviating firm obtains π^d; if both firms compete, they obtain π^n. We now add another player in the game, namely the competition authority. Its objective is to maximize consumer surplus and its instruments are the following: (i) it can impose fines on colluding firms and (ii) it can reduce fines (or give a reward) to firms or individuals bringing evidence of collusion.

Regarding evidence of collusion, we make the following assumptions. First, as discussed above, we realistically assume that collusion cannot occur without communication among the firms and that communication generates hard evidence (memos, reports of meetings, etc.). This evidence can be found by the competition authority if it audits the industry. We assume that audit takes place with probability ρ. The evidence can also be brought to the competition authority by a firm or by one of its employees (as in Cases 14.7 and 14.8). It is further assumed for simplicity that the evidence disappears at the end of the period. Let F denote the maximal fine imposed by the authority. In expectation, this fine is assumed to be too small to deter collusion, $\pi^c - \rho F > \pi^n$ (the expected net profit from collusion is larger than the profit from competition).

To take communication into account, we add an initial stage to the repeated game we considered in Section 14.2. In this initial stage, firms decide whether or not to communicate (about prices, quotas, punishments, etc.). If they both choose to communicate, communication takes place, evidence is generated and the repeated game ensues; that is, each firm chooses either to collude or to deviate and compete; each firm may in addition decide to report the evidence of collusion to the competition authority. In all other cases, that is, if at least one firm prefers not to communicate, there is no communication and thus no collusion; both firms have no other choice than to compete.

[39] This extension is due to Aubert, Rey and Kovacic (2006).

No revelation mechanism

As a benchmark, consider first the case where the competition authority can only rely on audits to detect collusion. The profits firms can obtain in that case are as follows. If at least one firm does not communicate, they both earn π^n. If both firms communicate, they both earn $\pi^c - \rho F$ when they both collude. If one firm deviates, it increases its profit to $\pi^d - \rho F$. Collusion can be sustained if the discounted profit from colluding is larger than the profit obtained when deviating, followed by eternal competition:[s]

$$\frac{1}{1-\delta}(\pi^c - \rho F) \geq (\pi^d - \rho F) + \frac{\delta}{1-\delta}\pi^n,$$

which is equivalent to

$$\frac{\delta}{1-\delta}[(\pi^c - \rho F) - \pi^n] \geq (\pi^d - \rho F) - (\pi^c - \rho F), \tag{14.18}$$

where the left-hand side is the discounted loss from punishment and the right-hand side is the immediate gain from deviation.

Leniency programme

Consider now a corporate leniency programme whereby a firm that reports evidence of collusion is granted a 'reward' R by the competition authority. We will discuss below whether this reward is a mere reduction of the maximal fine F ($R \leq 0$) or is actually a positive transfer ($R > 0$). As firms want to deter reporting, they will punish it in the same way they punish a deviation. As a result, if a firm decides to report, it will deviate as well. Clearly, denouncing is attractive only in the presence of some leniency; that is, the reward must be larger than the expected fine: $R > -\rho F$ (which is of course satisfied when the reward is positive). What changes with respect to the benchmark we just analysed is the profit for a reporting/deviating firm: it increases from $\pi^d - \rho F$ to $\pi^d + R$. The condition for the sustainability of collusion then becomes

$$\frac{\delta}{1-\delta}[(\pi^c - \rho F) - \pi^n] \geq (\pi^d + R) - (\pi^c - \rho F),$$

which is more stringent than (14.18) when $R > -\rho F$. Reversing the inequality, we can compute the minimum reward $R_{\min}$ that the competition authority must offer to induce a firm to report collusion:

$$R \geq R_{\min} = \frac{\delta}{1-\delta}[(\pi^c - \rho F) - \pi^n] - [\pi^d - (\pi^c - \rho F)].$$

If the minimal reward $R_{\min}$ is zero or negative, then leniency (whereby the reporting firm pays only a limited fine, or no fine at all) suffices to deter collusion. This is the case as long as the discount factor is not larger than $[\pi^d - (\pi^c - \rho F)]/(\pi^d - \pi^n)$. Otherwise, $R_{\min}$ is positive, meaning that the competition authority has actually to pay a reward to the reporting firm.

[s] Firms optimally use the grim trigger strategy. Recall also that evidence is assumed to disappear after one period; therefore, in the punishment phase that starts in the next period, firms earn π^n (and not $\pi^n - \rho F$).

Lesson 14.18 **Reducing fines for firms that report incriminating evidence may deter collusion. However, for some cartels, competition authorities have to grant sufficiently large, positive rewards to deter collusion.**

Several concerns may arise about the implementation of these, potentially large, rewards. First, large rewards may not be credible if the competition authority is budget-constrained; however, the fines paid by the other firms can be used to reward the informant. Second, public opinion may oppose the idea of granting rewards to guilty firms. This problem could be circumvented by secretly bargaining with the informant, but this would seriously undermine the predictability and credibility of the whole leniency programme. In practice, leniency programmes usually deny amnesty to ring-leaders. However, as the previous argument shows, offering rewards to any cartel member, including the instigator, improves collusion deterrence. A third issue is that large rewards could generate additional incentives to collude, for instance by inducing firms to 'take turns' for reporting collusion. To counter this, competition authorities grant leniency only to the first informants.

Whistleblowing programme

The corporate leniency programme can be adequately complemented by a programme that grants a positive reward, denoted B, to employees reporting incriminating evidence. Here, the idea is that firms will have to bribe informed employees to prevent them from disclosing information; this will make collusion less profitable and thus, harder to sustain. To show how useful such a programme can be, we consider a situation where full corporate leniency does not suffice to deter collusion, that is, $R_{\min} > 0$. Suppose that collusion requires that k employees be in the know, and that employees stay with the firm for one period only.[t] It follows that if a firm wants to prevent its employees from denouncing it would have to pay a total bribe of kB (in each period).

Assume first that the whistleblowing programme is the only instrument that the competition authority can use. In that case, a deviating firm is still exposed to prosecution and must compensate its informed employees. Hence, both the collusion and the deviating profits are reduced by the total bribe. It follows that collusion *cannot be sustained* if

$$\frac{\delta}{1-\delta}[(\pi^c - \rho F - kB) - \pi^n] < (\pi^d - \rho F - kB) - (\pi^c - \rho F - kB),$$

which is equivalent to

$$\begin{aligned} R_{\min} &= \frac{\delta}{1-\delta}[(\pi^c - \rho F) - \pi^n] - \left[\pi^d - (\pi^c - \rho F)\right] \\ &< \frac{\delta k B}{1-\delta} - \rho F. \end{aligned}$$

We observe that the latter condition becomes less stringent, meaning that collusion is easier to deter when the individual reward B, the number of informed employees k or the discount factor increase.

[t] The whistleblowing programme would even be more effective if employees stayed longer with the firm.

Assume now that the competition authority can combine corporate leniency and individual whistleblowing. If an informing firm can benefit from, say, full leniency, a deviating firm prefers reporting over bribing its informed employees; the profit of a deviating firm is thus π^d and the condition for collusion not to be sustainable becomes

$$\frac{\delta}{1-\delta}[(\pi^c - \rho F - kB) - \pi^n] < \pi^d - (\pi^c - \rho F - kB) \Leftrightarrow R_{\min} < \frac{kB}{1-\delta}.$$

Importantly, we observe that the competition authority may not manage to destroy collusion when it uses the leniency and whistleblowing programmes separately, but may succeed when it uses the two instruments in combination. This would occur when

$$\max\left\{\frac{\delta kB}{1-\delta} - \rho F, 0\right\} < R_{\min} < \frac{kB}{1-\delta}.$$

Lesson 14.19 **Corporate leniency and individual whistleblowing programmes are complementary in the fight against collusion.**

Review questions

1. Contrast the conditions for cartels to be stable when they form in a simultaneous versus a sequential way.
2. Explain how a collusive outcome may emerge non-cooperatively as the equilibrium of a repeated game. Discuss how the horizon of the game, the number of firms, the frequency of interaction and multimarket contact affect the sustainability of collusion.
3. Explain why tacit collusion is harder to sustain when demand fluctuates and when the rivals' actions cannot be observed.
4. Explain why competition authorities encourage colluding firms and their employees to report incriminating evidence through leniency and whistleblowing programmes.

Further reading

To compare the different procedures of cartel formation, we have followed the survey by Bloch (2002). A seminal analysis of cartel stability is due to d'Aspremont *et al.* (1983). Friedman (1971) was the first to model tacit collusion through 'supergames'. Abreu (1986, 1988) extended this analysis by examining optimal punishment schemes. The seminal analyses of tacit collusion under unobservable actions, demand fluctuations and multimarket contact are respectively due to Green and Porter (1984), Rotemberg and Saloner (1986) and Bernheim and Whinston (1990). Our analysis of leniency and whistleblowing programmes follows Aubert, Rey and Kovacic (2006). Recent surveys about the design of these programmes and about the detection of collusion are found, respectively, in Spagnolo (2008) and Harrington (2008).

Exercises

14.1 Collusion and quantity competition

Consider the following market. Two firms compete in quantities, that is, they are Cournot competitors. The firms produce at constant marginal cost equal to 20. The inverse demand curve in the market is given by $P(q) = 260 - q$. Suppose that the firms compete in this market for an infinite number of periods. The discount factor (per period) is δ, $\delta \in (0, 1)$.

1. Find the equilibrium quantities under Cournot competition as well as the quantity that a monopolist would produce. Calculate the equilibrium profits in Cournot duopoly and the monopoly profits.
2. The firms would like to collude in order to restrict the total quantity produced to the monopoly quantity. Write down grim trigger strategies that the firms could use to achieve this outcome.
3. For which values of δ is collusion sustainable using the strategies of (2)? (*Hint:* Think carefully about what the optimal deviation is.)

14.2 Collusion and pricing

Consider a homogeneous-product duopoly. The two firms in the market are assumed to have constant marginal costs of production equal to c. The two firms compete, possibly over an infinite time horizon. In each period they simultaneously set price p_i, $i = 1, 2$. After each period the market is closed down with probability $1 - \delta$.

Market demand $Q(p)$ is decreasing, where $p = \min\{p_1, p_2\}$. Suppose, furthermore, that the monopoly problem is well-defined, that is, there is a solution $p^M = \arg\max_p pQ(p)$. If firms set the same price, they share total demand with weight λ for firm 1 and $1 - \lambda$ for firm 2. Suppose that $\lambda \in [1/2, 1)$. Suppose that firms use trigger strategies and Nash punishment.

1. Suppose that $\delta = 0$. Derive the equilibrium of the game.
2. Suppose that $\delta > 0$ and $\lambda = 1/2$. Derive the condition according to which firm 1 and firm 2 do not find it profitable to deviate from the collusive price p^M.
3. Suppose that $\delta > 0$ and $\lambda > 1/2$. Derive the condition according to which p^M is played along the equilibrium path. Show that the condition is more stringent the higher λ.
4. Show that the previous results in (3) also hold for any collusive price $p^C \in (c, p^M)$.