Chapter 6

THEORIES OF OLIGOPOLY BEHAVIOR*

CARL SHAPIRO**

Princeton University

Contents

*This chapter is dedicated to the 150th anniversary of the publication of Cournot's (1838) theory of oligopoly.

**I thank the Institute for International Economic Studies at the University of Stockholm which provided an excellent environment while I was beginning to write this chapter, and the National Science Foundation, which provided financial support under Grant SES 86-06336. I also thank Hal Varian for his help in implementing this chapter on his VERTEX system, and Blaise Allaz for translating Bertrand's (1883) Book Review from French into English. Avinash Dixit provided extremely useful comments on a rough draft. Richard Gilbert, Gene Grossman, Barry Nalebuff, Rob Porter, Steve Salant, Jean Tirole, Hal Varian, and Robert Willig read a preliminary version and made numerous helpful suggestions.

Handbook of Industrial Organization, Volume I, Edited by R. Schmalensee and R.D. Willig

1. Introduction

The study of oligopolistic industries lies at the heart of the field of industrial organization. One's beliefs about the behavior of large firms in concentrated markets colors one's views on a broad range of antitrust and regulatory policies. And these beliefs derive in turn from theory and evidence of oligopolists' behavior. In this chapter, I present and evaluate the primary competing theories of oligopolistic behavior.

Oligopoly theory has a long history, as befits such a central topic in microeconomics. Sir Thomas Moore coined the term oligopoly in his *Utopia* (1516), and noted that prices need not fall to competitive levels simply due to the presence of more than a single supplier.[1] And we are about to mark the 150th anniversary of the publication of Cournot's (1838) pathbreaking book, wherein he provided the first formal theory of oligopoly. In an influential review of Cournot's book (it took economists some fifty years to become aware of Cournot's work) Bertrand (1883) criticized the theory developed by Cournot. Indeed, to read Edgeworth (1925), one would think that Cournot's theory had long since been discredited.[2] Despite this rather negative reception given to Cournot's theory,[3] it remains today the benchmark model of oligopoly. In fact, a glance at virtually any microeconomics textbook reveals that few if any subsequent developments in oligopoly theory are generally regarded as important enough for inclusion.[4] Part of the goal of this chapter is to clarify just how oligopoly theory has, and has not, progressed in the 150 years since Cournot developed his theory. Although oligopoly fits conceptually between the extremes of monopoly and perfect competition, its study requires a rather different set of tools, namely those of game theory.[5] The hallmark of oligopoly is the presence of strategic interactions among rival firms, a subject well suited for game-theoretic analysis. This chapter presumes a working knowledge of such concepts as Nash equilibrium and subgame

[1] I rely here on Schumpeter (1954, p. 305) for the reference to More.

[2] Writing in 1897, Edgeworth stated (1925, pp. 117–118) "Cournot's conclusion has been shown to be erroneous by Bertrand for the case in which there is no cost of production; by Professor Marshall for the case in which the cost follows the law of increasing returns; and by the present writer for the case in which the cost follows the law of diminishing returns."

[3] But see also Fisher (1898), who stated: "Cournot's treatment of this difficult problem [duopoly] is brilliant and suggestive, but not free from serious objections."

[4] The theories from the past century that most often are cited seem to be Stackelberg's (1934) modification of Cournot's theory and Sweezy's (1939) kinked demand curve theory.

[5] Actually, Cournot's plan was to smoothly fill in the cases between pure monopoly and perfect competition. Schumpeter notes, however (1954, p. 981), that "as we leave the case of pure monopoly, factors assert themselves that are absent in this case and vanish again as we approach pure competition," so that "the unbroken line from monopoly to competition is a treacherous guide".

perfect equilibrium, although the latter is developed somewhat in Sections 3 and 4.[6]

The material in this chapter is closely related to that in the chapters on collusion and cartels, entry deterrence, and product differentiation. In applications, these neighboring topics are inevitably intertwined with the analysis of noncooperative output, pricing, and investment policies that constitutes the bulk of the current chapter. For the purposes of this chapter, I define the boundaries of oligopoly theory by excluding explicitly cooperative behavior, by assuming a fixed number of active oligopolists,[7] and by avoiding any analysis of spatial competition.

Within these boundaries, we would like oligopoly theory to provide predictions regarding the relationship between market structure and performance, as measured, say, by price–cost margins. For a given set of active firms, and cost and demand conditions, what outputs and prices do we expect to prevail? Are two to three firms sufficient to ensure a relatively competitive outcome, or does this require many rivals? What institutional arrangements are conducive to (i.e. facilitate) relatively collusive behavior?

We also rely on oligopoly theory to provide a way of thinking about strategic behavior more generally, and its antitrust implications. By strategic behavior I mean not just production and pricing policies, but also decisions regarding investments, inventories, product choice, marketing, and distribution. Although there is no general theory encompassing this entire range of behavior, oligopoly theory can help us to understand many dimensions of strategic rivalry and to identify particular strategies that have anticompetitive effects.

Oligopoly theory has of course been the subject of numerous previous surveys. Stigler, in his influential 1964 paper, wrote (p. 44): "No one has the right, and few the ability, to lure economists into reading another article on oligopoly theory without some advance indication of its alleged contribution." Following this principle, it seems appropriate for me to explain the relationship between this chapter and previous surveys of oligopoly theory. Scherer (1980, chs. 5–7) is a standard reference at the undergraduate level. Dixit (1982) gives a short summary of three of the main strands of modern oligopoly theory. Friedman (1983) provides a fairly recent general coverage of the topic. Fudenberg and Tirole (1986a) is a selected survey of advanced topics in dynamic oligopoly, and

[6] I also assume that the reader is familiar with the concepts of pure and mixed strategies. See Chapter 5 by Fudenberg and Tirole in this Handbook for an introduction to the game-theoretic tools most often utilized in industrial organization.

[7] Here a firm is "active" if it has borne any sunk entry costs necessary to operate in the industry; thus I am steering clear of issues relating to entry deterrence. Throughout this chapter I also leave in the background the fundamental forces, such as economies of scale, that limit the number of firms in the industry. Of course, oligopoly theory is most interesting when the number of firms so determined is rather small. See Chapter 1 by John Panzar in this Handbook for a discussion of technology and market structure.

Kreps and Spence (1985) discuss many of the intriguing issues that arise in dynamic rivalry. But none of these authors makes an effort to report on and critique the vast number of articles on oligopoly in the past decade. The current chapter aims to be both systematic and comprehensive at a more advanced level than, say, Friedman (1983), but with an emphasis on the *economics* of oligopoly, rather than the game theory itself.

The various modern theories of oligopoly behavior are essentially a set of different games that have been analyzed; these games do not represent competing theories, but rather models relevant in different industries or circumstances. Accordingly, this chapter is organized on the basis of the structure of these games. Some readers may find that I give short shrift to such classic topics as price leadership or monopolistic competition. But this is a conscious choice in order to emphasize the more recent, game-theoretic contributions.

I begin in the next section with static models of oligopoly, distinguished according to the strategies available to the competing firms (quantities and prices being the two leading candidates). I then move on in Section 3 to games that are simple repetitions of the static games – finitely or infinitely repeated quantity or price games. Section 4 begins the study of strategic behavior by looking at two-period games. I discuss there many of the dimensions of competition that are absent in supergames. These include investments in physical capital, the establishment of a customer base, and R&D investments. These ideas are continued in Section 5, where I look at dynamic games in which the firms can make lasting commitments so that history matters – games that are *not* simple repetitions of the static competition. A summary and conclusion follow.

Before embarking on my analysis, it is best to provide the reader with a word of warning. Unlike perfect competition or pure monopoly, there is no single "theory of oligopoly". The rival theories presented below would seem each to have its appropriate application, and none can be considered *the* prevailing theory. Indeed, there has long been doubt about the wisdom of seeking a single, universal theory of oligopoly, and I share this doubt.[8] Only by making special assumptions about the oligopolistic environment – each of which will be appropriate in only a limited set of industries – can we expect to wind up with a specific prediction regarding oligopoly behavior. I view the development of oligopoly theory as providing us with an understanding of which environments lead to various types of equilibrium behavior, and with some sense of the methods by which large firms both compete and seek to avoid competition. But I do not expect oligopoly theory – at least at this stage of its development – to give tight inter-industry predictions regarding the extent of competition or collusion.

[8]Schumpeter (1954, p. 983), for example, appears to concur with Pigou's conclusion regarding the indeterminateness of duopoly equilibrium in his *Wealth and Welfare* (1912).

2. Static oligopoly theory

The natural starting place in a study of oligopoly theory, both logically and historically, is a static model of strategic interactions. Although a timeless model of oligopoly cannot, by definition, treat the essential issue of how rivals *react* to each other's actions, it does serve to elucidate the basic tension between competition and cooperation and provide an essential ingredient for the richer, dynamic analysis below. Of course, static oligopoly theory can only provide predictions about short-run behavior, taking the firms' capital stocks, and hence their variable cost functions, as given. I shall consider strategic investment plans in Sections 4 and 5 below.

Any theory of oligopoly must confront the essential tension in small-numbers rivalry: each firm is tempted to compete aggressively to increase its own market share, but if all firms do so, they all suffer. Another way of putting this is that oligopolistic interactions necessarily have the underlying structure of a prisoner's dilemma game.[9] What we can seek in a static theory of oligopoly is a prediction about the particular resolution of this tension between competition and cooperation, based upon such fundamentals as demand conditions, cost conditions, and the number of competing firms. As we shall see, even within the class of static models, the precise way in which rivalry is modeled has a profound impact on equilibrium behavior.

2.1. Cournot oligopoly: Competition in outputs

I present here a modern version of Cournot's (1838) theory of oligopoly. Consider n firms competing to supply a homogeneous good, the demand for which is given by $p(X)$, where p is the price, and $X \equiv x_1 + \cdots + x_n$ is industry output, x_i being firm i's output. Firm i produces according to the cost function $C_i(x_i)$. I shall at times denote firm i's marginal cost, $C_i'(x_i)$, by simply c_i; this is exactly accurate in the convenient case of constant marginal costs, but is otherwise a notational short-cut. Firm i's profits are

$$\pi_i = p(X)x_i - C_i(x_i)$$

if it produces x_i and total output is X.

[9] As discussed above, I restrict attention to noncooperative behavior; for a treatment of cartels, mergers, and explicit collusion, see Chapter 7 by Jacquemin and Slade in this Handbook.

2.1.1. The Cournot equilibrium

In a timeless model, each firm makes but a single decision, which in some way captures how "aggressive" is its attempt to make sales. In the Cournot model, all firms choose their outputs simultaneously. In other words, the Cournot equilibrium is a Nash equilibrium in quantities. Given a set of choices, $\{x_i\}$, price adjusts to clear the market, i.e. $p = p(X)$. Formally, the Cournot equilibrium output vector, $(x_1, \ldots, x_n)$, is determined by the n equations, $\partial\pi_i/\partial x_i = 0$, $i = 1, \ldots, n$. The ith equation typically is called firm i's *reaction curve*, since it represents firm i's optimal choice of x_i as a function of its rivals' choices.[10] Like any Nash equilibrium, the Cournot equilibrium is the set of self-enforcing actions from which no firm would unilaterally wish to deviate.

Maximizing π_i with respect to x_i, firm i's reaction curve is given by the first-order condition[11] $p(X) + x_i p'(X) = c_i$, which we can re-write as $p(X) - c_i = -x_i p'(X)$, or

$$\frac{p(X) - c_i}{p(X)} = \frac{s_i}{\varepsilon}, \quad i = 1, \ldots, n, \tag{1}$$

where $s_i \equiv x_i/X$ is firm i's market share, and $\varepsilon > 0$ is the market elasticity of demand at X, $\varepsilon \equiv -p(X)/Xp'(X)$. Equation (1) is the basic *Cournot oligopoly pricing formula*.

A Cournot equilibrium exists under quite general conditions, even for differentiated products. Basically, we need each firm's profits to be quasi-concave in its output. See Friedman (1971) for details. An overly strong condition that is sufficient for existence is that π_i actually be concave in x_i, i.e. $\partial^2\pi_i/\partial x_i^2 < 0$ at all x_i, X. This condition is of course just firm i's second-order condition when evaluated at the equilibrium point. In the case of homogeneous products, it becomes:

$$a_i \equiv 2p'(X) + x_i p''(X) - C_i''(x_i) < 0, \quad i = 1, \ldots, n. \tag{2}$$

If the demand function is concave, $p''(X) \leq 0$, and if the cost function exhibits nondecreasing marginal cost, $C_i'' \geq 0$, then equation (2) is satisfied.

More recently, Novshek (1985) has provided us with a somewhat weaker condition that nonetheless guarantees the existence of a Cournot equilibrium in

[10] For homogeneous goods, firm i's optimal quantity depends only upon its rivals' aggregate output, $X_{-i} \equiv X - x_i$.

[11] The second-order condition, $2p'(X) + x_i p''(X) - C_i''(x_i) < 0$, requires that the firm's marginal revenue curve intersect its marginal cost curve from above. I assume this condition is met for each firm. I also assume that the n firms each earn non-negative profits when their outputs are given by (1). In other words, I am avoiding issues of entry and exit, taking n as exogenous.

the case of homogeneous products. Assuming that the demand and cost functions are differentiable and monotonic, he shows that an equilibrium exists so long as $Xp'(X)$ declines with X, i.e. so long as

$$p'(X) + Xp''(X) \leq 0 \tag{3}$$

at all X. This condition is equivalent to

$$b_i \equiv \partial^2\pi_i/\partial x_i\,\partial x_j = p'(X) + x_i p''(X) \leq 0, \tag{4}$$

for all $x_i \leq X$.[12] Condition (4) states that firm i's marginal revenue must not rise with its rivals' outputs. Comparing equations (2) and (4) reveals that Novshek's generalization consists of greatly reducing the requirements on the cost functions that are needed to ensure existence.

Uniqueness of equilibrium is not nearly so general, as the reaction curves may easily intersect more than once. But Friedman (1977, p. 71 and p. 171) provides conditions under which equilibrium is unique. An overly strong sufficient condition for uniqueness is

$$\frac{\partial^2\pi_i}{\partial x_i^2} + \sum_{j\neq i}\left|\frac{\partial^2\pi_i}{\partial x_i\,\partial x_j}\right| < 0 \quad , i = 1,\ldots,n.$$

In the case of homogeneous goods, this uniqueness condition becomes $a_i + (n-1)|b_i| < 0$, $i = 1,\ldots,n$. Since a_i must be negative in equilibrium, this inequality can be rewritten as

$$|a_i| > (n-1)|b_i|, \quad i = 1,\ldots,n. \tag{5}$$

We shall see below that inequality (5) is useful when performing comparative statics exercises. Unfortunately, inequality (5), while sufficient for uniqueness, is far too strong: it usually is not met. In fact, direct substitutions demonstrate that (5) *must* be violated whenever (a) $p'' \leq 0$ and $n \geq 3$, or (b) inequality (3) holds and $n \geq 4$.

In the case of constant marginal costs, inequality (3) is sufficient for uniqueness. If this condition holds strictly, there can be at most one Cournot equilibrium in which all firms produce positive quantities. The proof of this fact is as follows. Suppose that two equilibria existed, with corresponding output vectors $(x_1^A,\ldots,x_n^A)$ and $(x_1^B,\ldots,x_n^B)$. Let $x^A \equiv \sum_{i=1}^n x_i^A$, and define x^B similarly. Label the equilibria so that $x^A \geq x^B$. This implies that the corresponding prices

[12] Clearly, (4) implies (3) by setting $x_i = X$. But (3) implies (4) as well: if $p''(X) \leq 0$, (4) is surely satisfied, and if $p''(X) > 0$, then (3) implies (4) since $Xp''(X) \geq x_i p''(X)$ for all $x_i \leq X$.

obey $p^A \leq p^B$. Comparing the first-order conditions for x_i, $p(X) + x_i p'(X) = c_i$, in each of the two equilibria, and using the fact that $p^A \leq p^B$, we have $x_i^A p'(X^A) > x_i^B p'(X^B)$ for all i. Adding up these inequalities across firms gives $x^A p'(X^A) > x^B p'(X^B)$. This last inequality is inconsistent with inequality (3) and the labeling convention $x^A \geq x^B$. Inequality (3) is therefore sufficient to rule out multiple interior equilibria.

2.1.2. *Characterization of the Cournot equilibrium*

The Cournot oligopoly pricing formula, equation (1), captures a number of features of oligopoly behavior that we might hope for from a theory of oligopoly. (a) Each firm recognizes that it possesses (limited) market power. There is a divergence between its price and marginal revenue: $MR_i = p(X) + x_i p'(X)$, so $p - MR_i = -x_i p'(X) > 0$. (b) The Cournot equilibrium is somewhere "in between" the competitive equilibrium and the monopoly solution. (c) The greater is the market elasticity of demand, the smaller are the markups at each firm. (d) The markup at firm i is directly proportional to that firm's market share. (e) The market shares of the firms are directly related to their efficiencies.[13] But (f) less efficient firms are able to survive in the industry with positive market shares.

In the symmetric case, all firms have the same cost function, and (1) becomes:

$$\frac{p-c}{p} = \frac{1}{n\varepsilon}, \tag{6}$$

where c is the common level of marginal cost. Of course, for $n = 1$, this equation is simply the monopoly markup formula, $(p - c)/p = 1/\varepsilon$. Equation (6) captures the notion that markets with more (equally placed) rivals perform more competitively. Indeed, as the number of firms grows, (6) indicates that prices approach marginal costs. But with some range of decreasing average costs, it is not possible for each of n firms to earn non-negative profits in a Cournot equilibrium if n is sufficiently large. So one must exercise considerable care before one can conclude that the Cournot equilibria approach the perfectly competitive equilibrium as n approaches infinity.[14]

[13] Note also that by summing (1) across firms one obtains an equation relating the equilibrium price to the sum of the firms' marginal costs (as well as n and ε). Cournot (1838, p. 86) was aware of this aggregation property.

[14] As Novshek (1980) points out, simply adding more firms cannot give the traditional Marshallian outcome: with fixed demand, as n grows each firm's Cournot output must approach zero, whereas with U-shaped average cost curves the competitive limit calls for strictly positive production levels at each firm. Instead, Novshek takes the limit of free-entry Cournot equilibria as the minimum efficient scale of operation becomes small in comparison with demand. With this limiting procedure, he establishes both that Cournot equilibria with free entry exist as $n \to \infty$ and that they approach perfect competition in the limit.

Quite generally, the Cournot equilibrium is not Pareto optimal from the point of view of the firms.[15] With each firm maximizing its own profits, given its rivals' outputs, the result cannot be maximal overall profits, since increases in a single firm's output have a (negative) effect on its rivals' profits. This "negative externality" causes the Cournot equilibrium to entail a higher aggregate output and lower price than does the collusive outcome. The inability of the firms to achieve the collusive outcome as a noncooperative equilibrium reflects the underlying prisoner's dilemma structure of the problem: each firm has an incentive to defect from collusion by producing more output, and all the oligopolists end up with lower profits due to these defections.

With quantity competition, firm i's optimal output typically is a decreasing function of its rivals' aggregate output. This will be the case as long as a given firm's marginal revenue is reduced when a rival increases its output. Formally, this is so if and only if $\partial^2\pi_i/\partial x_i\,\partial x_j < 0$, which is exactly condition (4) above, namely $b_i < 0$. Hahn (1962) assumed this condition to hold for all firms.[16] When the Hahn condition is met, the firms' *reaction functions are downward sloping*, a fact of considerable significance when we explore multistage models below.

Another way to characterize the Cournot equilibrium is to ask the following question: "What does a Cournot equilibrium maximize?" We know that the answer to this question is neither industry profits (since the Cournot equilibrium does not replicate collusion) nor social welfare (since prices are not equated to marginal costs). In fact, Bergstrom and Varian (1985a) have shown that the Cournot equilibrium maximizes a *mixture* of social welfare and profits.

Define gross benefits by the total area under the demand curve,

$$B(X) \equiv \int_0^X p(z)\,\mathrm{d}z,$$

and define total welfare as the sum of producer and consumer surplus, or, equivalently, total benefits less total costs,[17]

$$W(X) \equiv B(X) - nC(X/n).$$

Then Bergstrom and Varian show that the first-order conditions for the Cournot equilibrium are the same as would arise if a social planner were aiming to

[15] Of course, collusion, while Pareto optimal for the firms, leaves consumers even worse off than in the Cournot equilibrium. Neither the Cournot equilibrium nor the collusive outcome is Pareto optimal when consumers and firms are considered together.

[16] A sufficient condition for $b_i < 0$ is that $p''(X) < 0$ at all X, i.e. that demand be concave.

[17] Here I assume that the n firms are equally efficient, and restrict attention to symmetric equilibria.

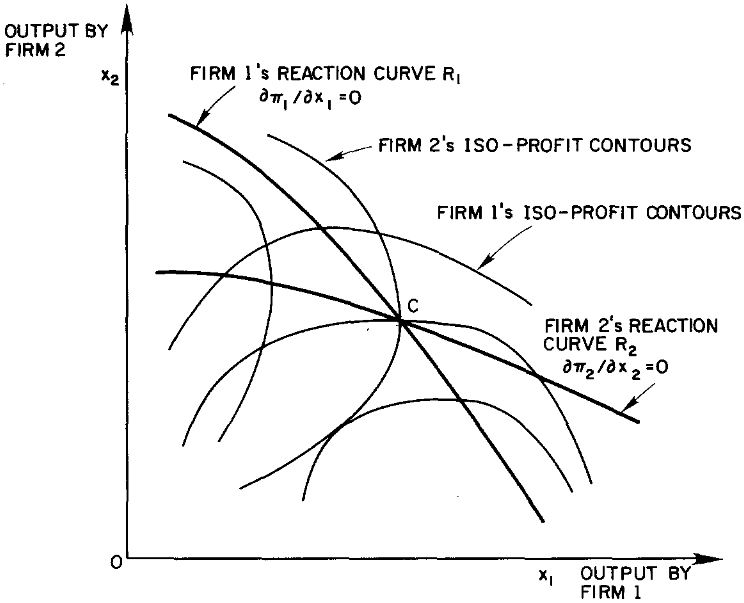

Figure 6.1. Cournot duopoly.

maximize the following function:

$$F(X) \equiv (n-1)W(X) + \pi(X),$$

where $\pi(X) \equiv p(X)X - nC(X/n)$ is industrywide profits. In other words, the Cournot equilibrium maximizes a weighted sum of welfare and profits. As the number of firms becomes large, the Cournot equilibrium becomes closer and closer to welfare maximization. Since welfare is the sum of consumer surplus, $S(X) \equiv B(X) - p(X)X$, and profits, $\pi(X)$, maximizing $F(X)$ is equivalent to maximizing $(n-1)S(X) + n\pi(X)$. In comparison with a social planner, extra weight is put on profits over consumer surplus, especially when n is small.

A diagrammatic treatment of Cournot duopoly is provided in Figure 6.1. Firm i's reaction curve is labeled R_i, $i = 1, 2$. The intersection of the two reaction curves at point C represents the Cournot equilibrium. Several of each firm's isoprofit contours, $\pi_i(x_1, x_2) = \bar{\pi}$, are included in the figure to show their relationship with the reaction curves. Tangencies between these isoprofit contours correspond to Pareto-optimal outcomes for the two firms. Observe that both firms' profits would increase if they could reduce their outputs in a coordinated fashion. Of course, by the definition of the Cournot equilibrium, neither firm *alone* could raise its profits with such a contraction. Typically, the joint-profit maximizing point lies to the southwest of the Cournot equilibrium point, as shown in Figure 6.1.

2.1.3. An historical note: The dual to Cournot equilibrium

Before leaving the topic of Cournot equilibrium, it is instructive to compare the standard Cournot oligopoly model, originally presented in Cournot's book in chapter 7, with his treatment of the "Mutual Relations of Producers" in chapter 9. In the latter chapter, he studies a situation in which two monopolists – one controlling copper, the other zinc – sell to the competitive brass industry. Brass is produced in fixed proportions from copper and zinc. Let the demand for brass be given by $D(p_{\mathrm{b}})$, where p_{b} is the price of brass. Scaling units so that one unit of brass is produced using one unit each of copper and zinc, we must have $p_{\mathrm{b}} = p_{\mathrm{c}} + p_{\mathrm{z}}$, where the prices of copper and zinc are p_{c} and p_{z}, respectively.

Cournot studies the equilibrium in the game where the copper and zinc monopolists each set prices simultaneously. Letting their marginal costs be c_{c} and c_{z}, and solving for the reaction functions $\partial\pi_{\mathrm{c}}/\partial p_{\mathrm{c}} = 0$ and $\partial\pi_{\mathrm{z}}/\partial p_{\mathrm{z}} = 0$, we find the equations analogous to equation (1):

$$\frac{p_i - c_i}{p_i} = \frac{1}{\sigma_i \varepsilon}, \quad i = \mathrm{c}, \mathrm{z}, \tag{7}$$

where $\sigma_i \equiv p_i/p_b$ is firm i's factor share and ε is the elasticity of demand for brass.

Equation (7) gives the *dual* to the usual Cournot equation, (1). Notice especially that each firm sets a *higher* price than would a coordinated monopolist controlling both copper and zinc, whose behavior would be described by the single equation $(p - c)/p = 1/\varepsilon$, where $p \equiv p_{\mathrm{c}} + p_{\mathrm{z}}$ is a composite price and $c \equiv c_{\mathrm{c}} + c_{\mathrm{z}}$. Here, each monopolist ignores the fact that raising his own price *reduces* the demand and profits for the other monopolist. With these "negative externalities", prices are set too high from the point of view of joint-profit maximization. Another way to see this is to recognize that each monopolist would have an incentive to raise his price above the coordinated monopoly level if the other were pricing at the level that maximizes joint profits.[18]

[18]This example generalizes easily to n inputs in much the same way as does the usual Cournot equilibrium. The finding of super-monopolistic pricing is reminiscent of the double markup problem in the case of a chain of vertical monopolies. It is interesting also to note that Cournot adopted a pricing game for his copper and zinc example, in contrast to his better-known quantity game in the conventional oligopoly context. See Bergstrom (1979) for a generalization of Cournot's copper and zinc example to many factors of production and neoclassical production functions, and for an integration of Cournot's two theories of duopoly, stressing the duality relationship between them. An exercise for the reader is to work out the Nash equilibrium in quantities in the copper and zinc example.

Cournot's copper and zinc example illustrates a general principle in oligopoly theory: results which apply when discussing competition among suppliers of *substitutes*, the usual subject of oligopoly, are typically reversed when considering rivalry among suppliers of *complementary* goods. See Singh and Vives (1984) and Vives (1985) for more recent examples of this principle.

2.1.4. Comparative statics

Often one is interested in the properties of Cournot equilibria as underlying parameters shift. As we shall see below, many of the principles in *multiperiod models* rely formally on the comparative statics properties of the basic Cournot model: in two-period models (see Section 4 below), for example, first-period actions alter the parameters underlying the second-period competition. My discussion here relies on Dixit's (1986) excellent and quite complete analysis of comparative statics for oligopoly.

For the purposes of comparative statics questions, it is important to determine if the reaction functions intersect in the "right way", i.e. so that Cournot's naive adjustment process (each firm adjusting its output in response to the other's according to its reaction curve) is stable. In the duopoly case, this requires that R_1 intersect R_2 from above, as shown in Figure 6.1. More generally, Dixit provides us with a simple set of sufficient conditions ensuring "stability". Suppose that each firm's second-order condition is met, i.e. that inequality (2), $a_i < 0$, is satisfied for all i, and that the Hahn condition, $b_i < 0$, holds. Then inequality (5), the condition ensuring uniqueness of equilibrium, is sufficient for stability (although it remains a very strong condition that often is violated – see the discussion above regarding uniqueness of equilibrium). Basically, both uniqueness and "stability" are ensured if the reaction curves only can intersect in the "right way", as described above.[19] In what follows, I assume that the appropriate "stability" conditions are in fact met. But the reader should be aware that the comparative statics results reported here, and the findings in Section 4 below, are reversed if these conditions are not met.

Consider first the effect of changing a parameter that affects only firm i's "marginal profitability", $\partial \pi_i / \partial x_i$. Such a shift might arise, for example, from a change in that firm's capital stock, R&D findings, or tax treatment. Consider a change of this type that is favorable to firm i in the sense of raising that firm's marginal profitability; a reduction in marginal costs or in taxes would have this effect. Then the prima facie effect of the change is to induce x_i to increase, since $\partial \pi_i / \partial x_i > 0$ at the original equilibrium point. In fact, after tracing through the

[19]See Dixit (1986, p. 117) for further discussion of (weaker) stability conditions, and such conditions in models with differentiated products or conjectural variations. See also Seade (1980) for an earlier discussion of stability and instability.

full equilibrium effect of such a change, it is possible, using standard comparative statics techniques, to show the following: (i) x_i will rise, (ii) X will rise, (iii) π_i will rise, and (iv) x_j and π_j will fall for all $j \neq i$.[20]

This type of shift favorable to firm i occurs in two-period models when firm i undertakes first-period investments in physical capital or in R & D that lower its own second-period marginal costs or shift out its second-period demand. Another example of this type of shift would arise if firm i were to enter into a separate market, such as a foreign market. Sales in the foreign market would lower the firm's marginal costs in the home market if marginal costs decline with output, $C_i'' < 0$, and raise them if $C_i'' > 0$. The comparative statics properties of the Cournot equilibrium in the home market provide information about the incentives to enter the foreign market (see Section 4 below for further discussion of this point).

Next, consider the effect of a shift in an industrywide parameter, say one that enhances demand, reduces all firms' costs, or reduces their taxes. Such a shift will increase the marginal profitability at *all* firms. Seade (1985) treats this case in some detail, focusing on the effects of an industrywide cost or tax increase (affecting all firms equally). In the case of constant marginal costs, Seade shows that the key parameter for evaluating comparative statics effects is the *elasticity of the slope of the inverse demand function*, $E \equiv -Xp''(X)/p'(X)$.[21] He shows that each firm's output will fall in response to the cost increase. More surprisingly, he finds that price will rise by *more* than marginal cost (e.g. by more than the size of an excise tax – a greater than full passthrough of the tax) if and only if E exceeds unity, as it must if demand is isoelastic. He also shows in the case of symmetric oligopoly that profits will increase with the cost or tax increase if and only if E exceeds 2, which it does for any constant elasticity demand curve with elasticity less than unity.[22] The reason for this perverse result is that the output contraction (and thus the higher price) induced by higher costs more than offsets the direct effect of these higher costs.

Finally, consider a shift in the *distribution* of costs across firms. If such a shift leaves the sum of the firms' (constant) marginal costs unchanged, it is in some sense orthogonal to the industrywide cost or tax shift just discussed. Since the total output X depends only on the sum of the firms' marginal costs, as noted earlier, such a shift will not alter price or aggregate output. With X constant as the c_i's vary, we can use firm i's first-order condition for x_i directly to see that

$$\Delta x_i = \frac{\Delta c_i}{p'(X)},$$

[20]Again, see Dixit (1986, pp. 120–121) for details.

[21]For the case of iso-elastic demand, $E = 1 + 1/\varepsilon$, where $\varepsilon > 0$ is the elasticity of demand.

[22]But note that the requirement of "stability" puts an upper bound on the value of E; if $n = 5$, for example, stability requires that $E < 6$.

so long as all firms continue to produce positive quantities. Thus, as Bergstrom and Varian (1985b) point out, each firm's shift in output is proportional to its shift in costs, with the constant of proportionality, $1/p'(X)$, equal across firms.

2.1.5. *Performance measures*

Another attractive feature of the Cournot equilibrium is that it allows us to draw some direct relationships between market structure and performance, where structure is captured via n and ε and performance is measured by the sum of consumer and producer surplus. These relationships are derived by aggregating equation (1) across firms.

A natural way to gauge the performance of an oligopolistic industry is to see how large are the firms' markups; we know that large divergences between price and marginal cost are related to poor performance and substantial market power. Define the *industrywide average markup* as the average of the firms' markups, weighted by their market shares:

$$\frac{p-\bar{c}}{p} = \sum_{i=1}^{n} s_i \frac{p-c_i}{p}.$$

Substituting for firm i's markup from (1), we have in a Cournot equilibrium:

$$\frac{p-\bar{c}}{p} = \sum_{i=1}^{n} \frac{s_i^2}{\varepsilon}$$

or

$$\frac{p-\bar{c}}{p} = \frac{H}{\varepsilon}, \tag{8}$$

where $H \equiv \sum_{i=1}^{n} s_i^2$ is the Herfindahl index of concentration.

Equation (8) suggests that, in the Cournot equilibrium, there is a negative relationship between the Herfindahl index and industry performance. It is only suggestive, however, unless we provide a solid foundation for using the industry-wide average markup as a welfare measure. Dansby and Willig (1979) show how to provide a solid welfare basis for a formula much like (8). They develop the theory of *industry performance gradient indexes*. Basically, beginning with any industry configuration of prices and outputs, their index is defined as the answer to the following question: "By how much would welfare rise if we could perturb the industry a small amount in the optimal welfare-improving direction?"[23] High values of the index correspond to poor performance, at least in the local sense that welfare would rise sharply if the firms could be induced (by, say, antitrust or regulatory policies) to expand their outputs slightly. For many policy purposes, small changes in the industry output vector are all that we can hope to induce.

The general expression for the industry performance gradient index, ϕ, is

$$\phi = \left(\sum_{i=1}^{n} \left(\frac{p_i - c_i}{p_i} \right)^2 \right)^{1/2}, \tag{9}$$

where the formula allows for heterogeneous products since the firms' prices, the p_i's, need not be equal. For the case of Cournot oligopoly with homogeneous products, $p_i = p$ for each i, and we can use equation (1) in conjunction with (9) to derive the industry performance gradient index as

$$\phi = \sqrt{H}/\varepsilon,$$

a modified version of equation (8). Thus, Cournot's theory provides an intuitively reasonable prediction of the relationship between equilibrium market structure (as measured by market shares), and performance. More concentrated industries have a higher ϕ, capturing poorer (local) performance. Amazingly, this relationship can be summarized using a simple concentration index, the Herfindahl Index, in conjunction with the elasticity of industry demand, ε.

2.2. Bertrand oligopoly: Competition in prices

A natural objection to the Cournot quantity model is that in practice businesses choose prices rather than quantities as their strategic variables. Indeed, the actual process of price formation in Cournot's theory is somewhat mysterious. Bertrand (1883), in his review of Cournot's book, was the first to criticize Cournot on these grounds, and his name has since been attached to simple pricing games, just as Cournot's is with simple quantity games. Stigler's (1964) excellent discussion of oligopoly theory makes a quite convincing case (see especially pp. 45–48) that one must pay attention to the particulars of price-cutting strategies, including selective discounts, if one is to develop a genuine understanding of oligopoly (although Stigler's interest is in the policing of tacit collusion and his arguments are explicitly dynamic).

[23]A "small amount" is captured by constraining the output perturbation vector to have unit length, where the prices are used as scaling factors so that the units of distance are dollars. See Dansby and Willig (1979, pp. 250–251).

Bertrand pointed out that with prices as strategic variables, each of two rival firms would have a strong incentive to undercut the other's price in order to capture the entire market. With equally efficient firms, constant marginal costs and homogeneous products, the only Nash equilibrium in prices, i.e. *Bertrand equilibrium*, is for each firm to price at marginal cost.[24]

To verify that the Bertrand equilibrium involves marginal cost pricing, one must simply check that neither firm would benefit from charging a different price, given that its rival prices at marginal cost.[25] To see that no other pricing pattern is a Nash equilibrium, label the firms so that $p_1 \leq p_2$ and consider any candidate equilibrium in which firm 1 sets a price above marginal cost. The equilibrium cannot involve firm 2 pricing strictly higher than firm 1, since firm 2 would then earn no sales and could increase its profits by undercutting firm 1 (slightly). Nor can the equilibrium involve firm 2 matching 1's price, as each firm would then have an incentive to (slightly) lower its price in order to capture all, rather than half, of the market.

Bertrand equilibria are equally easy to derive in the case where the firms' costs are unequal, so long as the assumption of constant returns to scale is retained. With n firms, if firm i has constant marginal costs of c_i, and we label the firms so that $c_1 < c_2 \leq \cdots \leq c_n$, then the Bertrand equilibrium involves firm 1 serving the entire market at a price of $p = c_2$, so long as c_2 does not exceed the monopoly price for a firm with unit cost c_1.[26] Here we see dominance by the most efficient producer, firm 1, who is partially disciplined by the presence of firm 2. Unlike the Cournot equilibrium, industry output is produced at least cost. Like the Cournot equilibrium – and in contrast to the simpler Bertrand equilibrium – the equilibrium is *not* the first-best allocation, since prices faced by consumers are in excess of marginal cost.

Bertrand equilibria are not without their own problems, especially with homogeneous goods. The greatest difficulty is that, with homogeneous goods (or with close but not perfect substitutes), Bertrand equilibria in pure strategies typically fail to exist absent the special assumption of constant marginal cost. In the central case of *increasing* returns to scale, "destructive competition" drives prices down to marginal cost, but this cannot be an equilibrium as prices then fail to cover average costs. Adding even a small fixed cost to the basic Bertrand model

[24]As is usually done, I assume here that a firm stands ready to serve all customers at its quoted price. Although not restrictive in the case of constant marginal costs, this assumption does matter when there are increasing marginal costs.

[25]The addition of more firms does not alter the Bertrand equilibrium.

[26]One must be careful to avoid the open set problem that arises because firm 1 would like to set its price as close to c_2 as possible but not actually at c_2, as that would permit firm 2 to share the market at $p = c_2$. Technically, one can avoid this problem by assuming that firm 2 declines to match the price c_2, since it is indifferent to doing so. I prefer to think of firm 1 as pricing "just below" c_2, and each other firm as being unable to gain sales at its best "credible" price, i.e. the lowest price at which it would not lose money were it to make sales.

of constant marginal costs causes nonexistence of equilibrium. Since oligopoly theory is most relevant is markets with significant scale economies, this lack of existence (or reliance on mixed strategies) must be considered a serious drawback to the application of Bertrand equilibria.[27]

Edgeworth (1925, pp. 118–120) provides an example of the nonexistence of Bertrand equilibrium in the presence of *decreasing* returns to scale. Edgeworth's nonexistence argument runs as follows. He assumes that each of two firms has constant marginal cost up to some capacity level. Suppose that neither of the duopolists has sufficient capacity to serve the entire market if price is at marginal cost, but each can accommodate more than half of the market at that price. As in the case of unlimited capacities, there can be no equilibrium with the firms charging equal prices above marginal cost, for each firm would have excess capacity and an incentive to undercut the other (slightly). Nor can equilibrium have one firm undercutting the other, for then the lower-priced firm should raise its price (retaining some discount). But, due to the capacity constraints, it also cannot be an equilibrium for each firm to set price at marginal cost. In such a configuration, either firm would have an incentive to *raise* its price: doing so would allow it to make *some* sales, since its rival cannot meet the entire industry demand at marginal cost. These sales would generate some revenues in excess of variable costs, and the profits so earned must exceed those from selling to half of the market at marginal cost. Since each firm has an incentive to undercut its rival's price when that price is high, but raise price when the rival is pricing at marginal cost, Edgeworth suggested that the market would fail to settle down, and rather that prices would *cycle* between high and low values. This theoretical pricing pattern is known as an Edgeworth cycle, although Edgeworth did not formally analyze the dynamic pricing game (see Section 5 below).

Although Bertrand equilibria in pure strategies do not generally exist in the absence of the constant returns to scale assumption, *mixed* strategy equilibria do exist under quite general conditions.[28] General mixed-strategy existence theorems for Bertrand–Edgeworth competition are available in Dasgupta and Maskin (1986), and are summarized nicely in Maskin (1986). See also Levitan and Shubik (1972). Kreps and Scheinkman (1983) look at Edgeworth's case of constant marginal costs up to some capacity level, and show how the nature of equilibrium depends upon the firms' capacities in relation to demand. Essentially, equilibrium may be of three possible types. If capacities are large, we get the original

[27]But note that with a reinterpretation of the timing of production and pricing, the nonexistence problem is solved in the contestability literature. The key there is that a firm need not actually incur any of the fixed costs unless it actually succeeds in capturing the market.

[28]All models of Bertrand–Edgeworth rivalry must specify a rationing rule which determines how a limited quantity available at a lower price is allocated among consumers. Such a rule determines the remaining (residual) demand facing the higher-priced firm. The existence of a mixed strategy equilibrium is not sensitive to the choice of rationing rule, although the equilibrium strategies are.

Bertrand equilibrium with prices at marginal cost. If capacities are small, each firm simply sets the price consistent with each producing to capacity. And for intermediate levels of capacity, we have a mixed strategy equilibrium.[29] The necessity of resorting to mixed strategies must, however, be considered a drawback associated with pricing games in homogeneous goods markets. And this technical necessity is related to a fundamentally unrealistic feature of such games: a firm's sales (and payoff) are a discontinuous function of its strategy (i.e. its price).[30]

Even in the special case of constant costs in which the Bertrand equilibrium does exist, it entails marginal cost pricing, independent of the number of firms or the elasticity of aggregate demand. This extreme prediction about the relationship between market structure and markups is not in accord with the bulk of the empirical evidence on oligopoly (see Chapter 17 by Tim Bresnahan in this Handbook). Both because its predictions are unrealistic, and because of existence problems, the Bertrand equilibrium has not become the standard static oligopoly theory. For homogeneous goods, Cournot's model remains the workhorse oligopoly theory.

Many of the difficulties with Bertrand equilibria are mitigated when the competing products are not perfect substitutes. With product differentiation, sales and profits are no longer discontinuous functions of prices.[31] And the Bertrand equilibrium involves prices above marginal costs, since each firm retains some market power by virtue of product heterogeneity.

To discuss differentiated-product pricing equilibria, we must first specify the demand system. Begin with the general demand functions (written in their direct form), $x_i = D_i(p_1, \ldots, p_n)$. Writing $\boldsymbol{p} = (p_1, \ldots, p_n)$, firms i's profits are given by

$$\pi_i = p_i D_i(\boldsymbol{p}) - C_i(D_i(\boldsymbol{p})). \tag{10}$$

The n equations or reaction functions characterizing the Bertrand equilibrium are of course $\partial\pi_i/\partial p_i = 0$, $i = 1, \ldots, n$. Using equation (10), firm i's reaction func-

[29]For demand $p = A - X$, constant marginal costs of c, and n firms with capacity $\bar{x}$ each, prices fall to c iff $\bar{x} \geq (A - c)/(n - 1)$, whereas the firms sell at their capacities iff $\bar{x} \leq (A - c)/(n + 1)$. In between we get the mixed strategy equilibrium. See Brock and Scheinkman (1985, proposition 1), for a succinct summary of this symmetric case.

[30]Note also that in a mixed strategy equilibrium each firm would have an incentive to change its ex ante optimal but ex post suboptimal price.

[31]Although product differentiation is clearly an important element of many oligopolistic industries, space does not permit me to treat it more systematically or in any depth. For a much more complete discussion of product differentiation, with emphasis on spatial competition, see Chapter 12 by Eaton and Lipsey, in this Handbook.

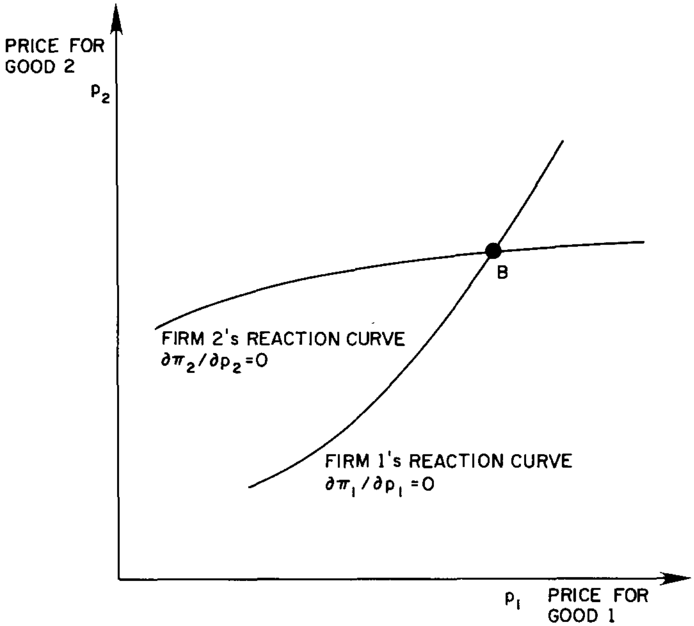

Figure 6.2. Bertrand duopoly with differentiated products.

tion can be written as

$$\frac{\partial \pi_i}{\partial p_i} = D_i(p) + (p_i - c_i)\frac{\partial D_i}{\partial p_i} = 0. \tag{11}$$

Of course, the existence of a Bertrand equilibrium is far from assured. We have already noted the nonexistence problem in the case of perfect substitutes, a problem that persists if the products are close substitutes. Here I characterize Bertrand equilibria with differentiated products without explicitly checking for existence.

The firms' reaction functions under pricing competition typically slope *upwards*, in contrast to those in Cournot equilibrium. This pattern is displayed in Figure 6.2. Firm i's optimal price is an increasing function of firm j's price if and only if $\partial^2\pi_i/\partial p_i\,\partial p_j > 0$. This condition is equivalent to

$$(p_i - c_i)\frac{\partial^2 D_i}{\partial p_i\,\partial p_j} + \frac{\partial D_i}{\partial p_j} - c_i''\left(\frac{\partial D_i}{\partial p_i}\right)\left(\frac{\partial D_i}{\partial p_j}\right) > 0. \tag{12}$$

Equation (12) is the dual of the Hahn condition, $\partial^2\pi_i/\partial x_i\,\partial x_j < 0$. Since $\partial D_i/\partial p_j > 0$ for substitutes, (12) is satisfied for a linear system, and in fact for any system with convex demand and cost functions, i.e. $\partial^2 D_i/\partial p_i\,\partial p_j > 0$ and $c_i'' > 0$.

With n firms, it is difficult to say much more about differentiated-product pricing equilibria without further assumptions about the demand system.[32] It is often helpful to assume a symmetric demand system. One convenient demand system[33] is derived from a benefit function:

$$B = G\left(\sum \varphi(x_i)\right), \tag{13}$$

which generates inverse demands of the form:

$$p_i = G'\left(\sum \varphi(x_i)\right)\varphi'(x_i). \tag{14}$$

By taking $G(S) \equiv S^\beta$ and $\varphi(x) \equiv x^\alpha$, one can study both the degree of substitutability between products and the demand for the entire product class by varying the parameters α and β. Another approach is to restrict attention to duopoly but preserve a more general demand system of the form $D_i(p_1, p_2)$, $i = 1,2$. For the purposes of working out an example, the simplest route is to explore duopoly with a special functional form. Perhaps the simplest demand system is the linear system described in inverse form by

$$p_i = \alpha_i - \beta_i x_i - \gamma x_j, \quad i, j = 1,2,\ i \neq j. \tag{15}$$

Here $\gamma > 0$ captures the notion of substitutes; $\gamma^2 = \beta_1\beta_2$ corresponds to perfect substitutes. With this linear demand system, the reaction curves also are linear, it is a simple matter to invert the system to write the demand function in their direct form, and both the Bertrand and Cournot equilibria can be directly computed.

2.3. *Cournot vs. Bertrand*

Which of these two competing static theories of oligopoly is "correct"? As a prelude to discussion of this question, it is instructive to see why the Nash equilibria in quantities and in prices are so different. The reason is that a single firm faces a very different *firm-specific demand* in the two cases. In duopoly, for

[32] Of course, one may derive the demand system directly from preferences over product characteristics. See Chapter 12 by Eaton and Lipsey in this Handbook for an analysis of such spatial competition.

[33] See Spence (1976) for further discussion of this demand system.

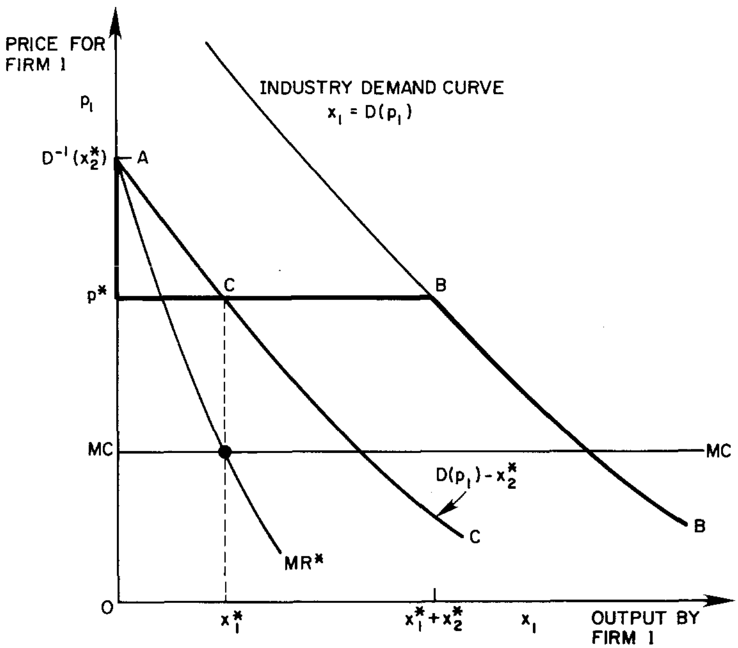

Figure 6.3. Quantity vs. price competition.

example, firm 1's demand as a Cournot duopolist is the locus ACC in Figure 6.3 (for $x_2 = x_2^*$). If, instead, the firms played a pricing game and firm 2 set a price $p^* = p(x_1^* + x_2^*)$, the Cournot price, then the demand facing firm 1 would be Ap^*BB. Firm 1 sees a *more elastic demand in the case of pricing competition*, and equilibrium prices are correspondingly lower. With quantity competition, each firm realizes that the other is committed to producing its announced quantity; with pricing competition, in contrast, each firm recognizes that it can take the entire market from its rival if it offers a lower price. This awareness leads to more aggressive behavior in the case of pricing competition.

Since the Bertrand equilibrium is so special, and in fact may not exist, in the case of homogeneous goods, it is instructive to extend our comparison of Cournot and Bertrand to the case of differentiated products. In fact, even in differentiated-product oligopoly models, Bertrand equilibria tend to be more competitive than are Cournot equilibria (although the sensitivity of equilibrium to the specification of the strategic variable is not quite so striking absent the homogeneity assumption). Under quite general conditions, pricing competition between duopolists leads to lower prices, higher outputs, and hence better performance than does quantity competition (for a given set of demand and cost functions). Singh and Vives (1984) provide a thorough comparison of Bertrand and Cournot equilibria for the special case of constant marginal costs and the linear demand system of equation (15). So long as the goods are substitutes, i.e. $\gamma > 0$, the Bertrand equilibrium is more competitive. Cheng (1985) provides a geometric proof of this same result that applies to a more general class of cost and demand functions. Vives (1985) gives fairly general conditions under which Cournot equilibria involve higher prices and profits (and lower welfare) than do Bertrand equilibria. The key to these comparisons is the difference in firm-specific demand depending upon the strategic variable being employed.

Apart from comparing price and quantity equilibria, we may wonder which one can be expected to arise, and in which industries. One approach to this question is to endogenize the choice of quantity vs. price strategies. Assuming that the technology of production and marketing makes either strategy feasible (pricing competition with unlimited capacities would *not* seem feasible if production is invariably subject to capacity constraints), it does seem fruitful to treat the choice of strategy as itself a variable. In their linear example, Singh and Vives (1984) consider a two-stage game in which each firm first selects "price" or "quantity" as its strategic mode, and then chooses a price or a quantity, according to its mode of operation. Singh and Vives show that it is a dominant strategy for each firm to select a quantity strategy, leading ultimately to the Cournot equilibrium. Each firm is better off if its rival faces a relatively inelastic demand, so that the rival will set a relatively high price or low quantity. This can be accomplished by adopting a quantity strategy.[34]

Klemperer and Meyer (1985) take a similar approach, but within a single-stage game where each firm may select a price or a quantity. They point out that, given its rival's choice, a firm can achieve any point on its residual demand curve equally well with a price or a quantity, so there are in general multiple equilibria. Both the Bertrand and the Cournot outcomes are equilibria in their model.[35] Klemperer and Meyer's main point, however, is that this multiplicity vanishes when uncertainty (about demand, say) is introduced. With uncertainty, a firm (monopolist or oligopolist) is in general *not* indifferent between a price and a quantity strategy. These authors go on to explore the factors – such as the slope of the marginal cost curves or the nature of the demand uncertainty – that determine whether the equilibrium entails price or quantity strategies.

An entirely different resolution of the "Cournot vs. Bertrand" debate is suggested by Kreps and Scheinkman (1983), although not within the context of a static model. They take the view, to which I adhere, that capital is a relatively

[34]But it is fairly unclear what it means for a firm to commit itself to a price or quantity strategy.

[35]This is an example of the multiplicity of equilibria when firms can select supply functions; see below for a discussion of more general supply-function equilibria.

sluggish variable, whereas prices can be adjusted rapidly. This leads to a model of capacity competition followed by pricing competition. In other words, firms first select capacities and then they play a pricing game subject to the endogenous capacity constraints inherited from the first period.[36] Kreps and Scheinkman show that the Cournot outcome is the equilibrium in this two-stage game, if one assumes a particular rationing rule for the second-period Bertrand-Edgeworth competition. They postulate the following rationing rule when firm 1 sets a lower price, $p_1 < p_2$, but can only serve demand $x_1 < D(p_1)$ at that price: the demand remaining for firm 2 is given by $D(p_2|p_1, x_1) \equiv D(p_2) - x_1$. This residual demand would arise if the consumers who place the *highest* value on having the product are the ones who purchase from firm 1.[37]

Kreps and Scheinkman's model is one way of dealing both with a major criticism of Cournot (it lacks a mechanism by which prices are set), and of Bertrand (with unlimited capacities it gives unrealistic predictions). But Davidson and Deneckere (1986) argue that Kreps and Scheinkman's finding is not robust with respect to the rationing specification. They point out that Kreps and Scheinkman's rationing rule is extreme in the sense that it leaves the *worst* possible residual demand for the higher-priced firm 2. Davidson and Deneckere explore the model of capacity choice followed by pricing competition using an alternative rationing rule that generates the *best* possible residual demand for the higher-priced firm. The alternative rule [which they attribute to Shubik (1959) and Beckmann (1965)] postulates that demand at the low price of p_1 is made up of a random sample of all consumers' demands at p_1. With this rationing rule, the residual demand is given by $D(p_2|p_1, x_1) \equiv D(p_2)[1 - x_1/D(p_1)]$, for all $p_2 > p_1$. Presenting numerical examples, Davidson and Deneckere show that, with this random rationing rule, equilibrium tends to be more competitive than under Cournot behavior. Their analysis further suggests that the static Cournot outcome cannot be supported as an equilibrium in the two-stage capacity and pricing game for most "intermediate" rationing rules.[38]

The choice between a pricing game and a quantity game cannot be made on a priori grounds. Rather, one must fashion theory in a particular industry to reflect the technology of production and exchange in that industry. For example, competition via sealed bids between firms without capacity constraints fits the Bertrand model quite nicely, whereas competition to install sunk productive capacity corresponds to Cournot.

[36]See Osborne and Pitchik (1984) for further analysis of this model, and Vives (1986) for a related model with a more general cost structure.

[37]Alternatively, it would arise if all consumers first have the opportunity to buy from firm 1 and are assumed to purchase according to their individual demands at that price, with each consumer retaining his individual residual demand for firm 2.

[38]See Dixon (1986) for further analysis of two-stage games as they relate to the Bertrand vs. Cournot debate.

After matching one's model of oligopolistic behavior to these features of production and distribution, one must test the theory's predictions against actual industry behavior. A common view is that pricing competition more accurately reflects actual behavior, but the predictions of Cournot's theory are closer to matching the evidence. Whichever static theory one prefers, it is clear that a serious theory of oligopoly behavior cannot be timeless.[39]

The Bertrand equilibrium concept suffers especially if one restricts attention to static models. The assumption implicit in the Bertrand model, that one firm can capture *all* of its rival's sales simply by offering a lower price, lacks realism. It is exactly in such a situation that the rival could be expected to respond most rapidly and vigorously, but reactions are ruled out entirely in this static theory. Within the context of static models, it seems less objectionable to assume, à la Cournot, that the rival is committed to a level of sales, such as would occur for a perishable good produced in advance or produced at low marginal cost subject to capacity constraints. Certainly, in most contexts prices can be changed quite rapidly; this need not be the case for quantities.

2.4. *Can there be reactions in a static model?*

In contrast to the entirely static view of Cournot equilibrium embodied in the statement that "the Cournot equilibrium is a Nash equilibrium in outputs", Cournot himself (1838, p. 81) thought of equilibrium being reached after an alternating sequence of quantity choices (reactions) by each of the duopolists, with each firm behaving according to its own reaction function.[40] This interpretation of the Cournot equilibrium is strained at best, however. Most simply, we are analyzing a static model, in which reactions are quite impossible. The alternating move game underlying Cournot's dynamics leads to a different outcome, since the Cournot reactions are myopic, not fully optimal, in that model (see Section 5 below). So, despite its widespread use, the term "reaction curve" is a misnomer.[41]

The unfortunate use of the term "reaction curve" in the study of Cournot oligopoly has contributed to a persistent confusion about the proper interpretation of the basic Cournot equilibrium. The correct interpretation is that the Cournot equilibrium point, like any Nash equilibrium point in a game, is a self-enforcing or self-confirming set of actions, from which no firm would want to

[39]This line of argument suggests that the most appropriate games to study are dynamic ones with price-setting. I report on the price-setting supergame literature below.

[40]As noted above, the convergence of this myopic adjustment process to the Cournot equilibrium point remains an essential "stability" condition when performing comparative statics or looking at multistage games.

[41]Dixit (1986, p. 110) observes that the phrase "firm i's equilibrium locus" would be superior to "reaction curve" but concedes that it is too late now to change our terminology.

deviate unilaterally. Instead of restricting their attention to the equilibrium point at which the reaction curves intersect, however, quite a few authors have attempted to interpret the reaction curves in a literal sense as responses to rival output choices. This effort has led to the study of the (pseudo) dynamics of the Cournot system, assuming that each firm responds to the other(s) according to its own reaction function. This line of research is a natural attempt to introduce responses into Cournot's model. And it is plain that rival responses are crucial in oligopoly. But one cannot properly analyze dynamic oligopoly using the simple Cournot reaction curves, as they are unlikely indeed to represent the firms' optimal dynamic responses.[42] A proper treatment of such dynamic games is reserved for Sections 3 and 5 below.

Another line of research attempts to retain a completely static structure but nonetheless generalize the Cournot model to include reactions. This approach represents an attempt to capture the concept of conscious parallelism – the recognition among oligopolists that price changes will likely be matched by rivals – in a static model. Despite the extreme conceptual problems of analyzing reactions within a static game, this approach has generated a considerable literature of its own. One reason for the hardiness of this questionable line of research is that proper dynamic models (see below) are much more difficult to work with, and lack the definitive results to be found in the static games.

An early example of the static approach to tacit collusion is the well-known *kinked demand curve* theory of oligopoly developed by Sweezy (1939) and Hall and Hitch (1939). Sweezy suggested that each firm would expect its rivals to match any price reductions, but not price increases. Given these expectations, price reductions provide no gain in market share, and are unprofitable. Similarly, a price increase causes a drastic loss of business. The conclusion is that the initial price – arbitrary though it was – is "stable", i.e. an equilibrium. Each firm perceives a demand curve with a *kink* at the initial price; demand is more elastic for price increases than for price decreases. The kink leads to a certain price rigidity, as modest shifts in demand or cost conditions will elicit no change in prices.[43] Plausible and popular as Sweezy's theory is, it strains the confines of the static model. Surely the speed of rival responses matters, and one should check that the postulated responses are in fact optimal.

The most common method of studying "reactions" in the static, homogeneous good model is to use the concept of *conjectural variations*, as developed by

[42]See Friedman (1977, p. 76) for references to the literature on Cournot pseudo-dynamics. As Friedman notes, "All the work is characterized by the assumption of single-period profit maximization, and it does not face up to the inconsistency stemming from the firm's incorrect assumptions about rivals' behavior. These shortcomings greatly reduce the potential value of the articles." I shall say a bit more about the literature on reaction functions in repeated oligopoly games in Section 3 below.

[43]But see Stigler (1947) for an empirical study undermining the kinked demand curve theory.

Bowley (1924). A firm's conjectural variation is defined as the response it *conjectures* about rival outputs if the initial firm alters its own output. Formally, firm i conjectures that $\partial X/\partial x_i = 1 + \nu_i$, where $\nu_i \equiv [\partial X_{-i}/\partial x_i]^e$ is firm i's *expectation*, or conjecture, about rivals' responses. With this conjecture, firm i perceives that an increase in its output will affect profits according to $\partial \pi_i/\partial x_i = p(X) - c_i + x_i(1 + \nu_i)p'(X)$, and is led via profit maximization to the modified reaction curve:

$$\frac{p(X) - c_i}{p(X)} = \frac{s_i(1 + \nu_i)}{\varepsilon}, \quad i = 1, \ldots, n. \tag{16}$$

In the Cournot model, $\nu_i = 0$ for every firm i, and equation (16) reduces to the usual Cournot equation, (1). A large value of ν_i captures a belief that other firms will "respond" aggressively to any attempt by firm i to increase its output. Such beliefs lead firm i to *less* aggressive behavior, i.e. a smaller output and a higher price, than under Cournot, as can be seen from (16). The expectation of aggressively competitive rival responses actually leads to more collusive equilibrium behavior. This apparent paradox will appear repeatedly in the supergame literature below. In terms of Figure 6.1, a higher value of ν_i shifts firm i's reaction curve inward.

In the symmetric case, a value of $\nu_i = n - 1$ captures a belief by firm i that its rivals will increase output in lock step with its own expansion, so firm i believes that it cannot increase its market share by increasing output. With such conjectures, each firm finds it optimal to produce its share of the collusive industry output; equation (16) becomes simply the monopoly markup formula, $(p - c)/p = 1/\varepsilon$. The other extreme value of ν_i is -1. In this case, firm i believes that $\partial X/\partial x_i = 0$, so $\mathrm{d}p/\mathrm{d}x_i = 0$: the firm is a price-taker, and sets price at marginal cost.[44]

Conjectural variations are a convenient way of parameterizing oligopolistic behavior. As such, they can be quite useful for comparative statics purposes. By parametrically changing the ν_i's, one can study the equilibrium of an oligopoly model for various behavioral parameters, i.e. various conjectures. But the idea behind conjectural variations is logically flawed, and they cannot constitute a bona fide theory of oligopoly. This is a fortiori true of the so-called "consistent conjectures" [Bresnahan (1981), Perry (1982)] which impose the requirement that, in the neighborhood of the equilibrium, a firm's conjecture about a rival's response equal the slope of that rival's reaction curve. Conjectural variations in general are an attempt to capture a dynamic concept, response, in a static model.[45] They may be useful if they can be shown to correspond to the equilibria

[44] A common error is to equate this conjectural variation outcome with the Bertrand equilibrium. The two are indeed equal if the firms have constant and equal marginal costs, but not otherwise.

[45] See also Makowski (1983) and Daughety (1985) for rather different attacks on the logical foundations of consistent conjectures. At another level, conjectural variations suffer in that they are inherently difficult, if not impossible, to measure.

in some class of dynamic oligopoly games, but at this time remain an ad hoc shortcut to the study of oligopolistic interactions.

An alternative way to build "reactions" into a static model is to specify strategies that are by their nature contingent upon rival behavior. For example, a strategy that consists of a *supply function*, $x_i(p)$, has this quality: the actual amount brought to market will depend upon rivals' announcements of their own supply schedules, since the price itself is jointly determined by all of the $x_i(\cdot)$'s. In fact, both Cournot and Bertrand strategies are special cases of supply functions. Cournot requires that a firm announce a perfectly *inelastic* supply function, and Bertrand insists that the supply function be perfectly *elastic*. Grossman (1981) presents an analysis of supply function equilibria in a market with large fixed costs and more potential producers than can be accommodated in the market. Under these conditions, supply function equilibria are extremely competitive.[46] Unfortunately, when all of the (potential) firms can be profitably accommodated in the market, as is the standard working assumption in oligopoly theory (ignoring issues of entry deterrence), there is a great multiplicity of supply function equilibria. The difficulty is that each firm is, given its rivals' supply functions, indifferent between a whole class of supply functions that yield the same equilibrium price and output. This indifference means that the firm could select supply functions specifying very different out-of-equilibrium behavior, and hence support any outcome.[47] What is missing is a notion of a "credible" or "reasonable" supply function for a firm to announce. Absent some restriction on the set of supply functions that firms can specify, there is no way to narrow down the set of equilibria in a useful fashion so as to yield further insights or theoretical predictions.

Another way to incorporate contingent strategies is to study equilibria in "contracts", or contingent price and quantity offers. Salop (1986) presents several illuminating examples demonstrating how contracts that are apparently quite competitive can serve instead as *facilitating practices*, promoting collusive outcomes.[48] One natural and interesting class of contracts discussed by Salop is that

[46] To achieve this result, Grossman restricts the set of allowable supply functions to those that would not lose money at any price. This is essentially a way of imposing "credibility" on the strategies without actually developing a dynamic model.

[47] Indeed, any output vector that allows each firm to earn non-negative profits can be supported as a supply function equilibrium. All that is required is that each firm be optimizing against its rivals' supply functions. Formally, these supply functions serve the same role as do conjectures in the conjectural variations literature, where again any individually rational point can be supported as an equilibrium with the appropriately chosen conjectures [see Laitner (1980) for a proof of this claim]. But Klemperer and Meyer (1986) suggest a possible resolution to this problem by introducing uncertainty about demand.

[48] Of course, many contracts among competing oligopolists can facilitate collusion. A merger (or profit-sharing contract) is the most obvious example; having each oligopolist sell his interest to a common third party is another possibility [see Bernheim and Whinston (1985) for a formalization of this using a common marketing agent]. I leave these contracts for Jacquemin and Slade to discuss in Chapter 7 of this Handbook on mergers and cartels, focusing attention here on the role of simple pricing contracts in facilitating collusion.

of *meeting competition clauses*. These provisions are contingent strategies in which a firm announces that its price is the minimum of some posted price, p, and the lowest price posted by another firm. In general, it is a Nash equilibrium for each firm to announce that its price is the minimum of the monopoly price and the lowest rival price. With these strategies, the monopoly price prevails, since each firm realizes that it cannot gain market share with a price reduction. Meeting competition clauses automatically incorporate – without any explicit dynamic structure – the aggressive responses to price-cutting that are needed to support collusion.[49] Put differently, meeting competition clauses formalize a particular set of supply functions or reactions that yield collusive outcomes.

Salop also analyzes *most favored nation clauses*, whereby a firm promises a given buyer that no other buyer will receive a better price. Again, an apparently pro-competitive clause may actually stifle competition: each firm is unable to engage in discriminatory price-cutting, and any type of secret price-cutting is more difficult to accomplish. Although each buyer may accept – indeed, value – the most favored nation clause, in equilibrium all buyers may pay higher prices as a result. A more complete analysis of games with these types of contingent strategies remains to be undertaken.[50]

All of the efforts described in this subsection are aimed at introducing reactions into the standard, static models of oligopoly. While the goal of this research is laudable, and indeed essential if one is to come up with an acceptable theory of oligopoly, the methods – especially those employed in the conjectural variations literature – are generally inappropriate. To study reactions, retaliation, price wars, and tacit collusion we require explicitly dynamic models of oligopoly.

3. Repeated oligopoly games

The limitations of static oligopoly have been evident at least since Stigler's (1964) classic paper. Stigler identified and stressed the importance of such factors as the speed with which competitors learn of a rival's price cut, the probability that such a competitive move is in fact detected, and the scope of retaliation by the other oligopolists.[51] Stigler's view of oligopoly theory as a problem of policing a tacitly

[49] But note that other equilibria also can be supported with meeting competition clauses. For any price between marginal cost and the monopoly price, it is an equilibrium for each firm to adopt a strategy of offering goods at that price and meeting any lower prices. Again we face the problem of multiple equilibria in contingent pricing strategies.

[50] Doyle (1985) is one effort along these lines that incorporates advertising into the game. Cooper (1986) shows how retroactive most-favored customer clauses can dampen firms' incentives to compete for customers in the second period of a two-period model; Holt and Scheffman (1987) undertake a related analysis. See Section 4 below for a further discussion of the effects of contracts on competition.

[51] Stigler also emphasized that tacitly colluding oligopolists seek a price structure involving some price discrimination, and that loyalty on the part of buyers is a key factor in assessing the profitability of any secret price-cutting. But these latter ideas have not been developed in the subsequent literature.

collusive industry configuration is now the norm. In the years since Stigler's paper was published, a great deal of work has been done to develop the theory of tacit collusion, particularly the role of defections and the reactions to them.

Quite generally, the success of oligopolists in supporting a tacitly collusive scheme depends upon their ability to credibly punish any defector from the scheme. Stronger, swifter, or more certain punishments allow the firms to support a more collusive equilibrium outcome, just as a higher value of the conjectural variation parameter, v_i, led firm i to behave more collusively in Subsection 2.4 above. This relationship tends to create some peculiar results: anything (such as unlimited capacities) that makes more *competitive* behavior feasible or credible actually promotes *collusion*. I call this the *topsy-turvy* principle of tacit collusion. The very competitive behavior is reserved as a threat to punish those who undermine the tacit collusion; it is never actually invoked in equilibrium, at least in nonstochastic models, which involve no "mistakes" or price wars in equilibrium.

Perhaps the most natural – and seemingly simple – way to go beyond static oligopoly theory is to study *stationary environments* in which each firm repeatedly sets its price or output and can respond to such choices by its rivals. Such models are called *repeated games*, and those with infinite repetitions are denoted *supergames*. The game played at each date is called the *stage game*. Repeated games are a formal way of focusing attention on reactions, since they involve play over time but the underlying environment is unchanging and the firms are unable to make lasting investments or commitments.[52] The past has no tangible effect on the present or the future; any effect arises purely because the oligopolists *remember* what has happened and condition their actions on that history.

3.1. Finitely repeated oligopoly games

3.1.1. Perfect equilibria in finitely repeated oligopoly

The simplest possible repeated oligopoly game involves a two-fold repetition of, say, Cournot oligopoly. At least in principle, this game permits each firm to respond (in the second period) to its rivals' actions (in the first period). Indeed, looking at Nash equilibria (as opposed to subgame perfect equilibria) in the twice-repeated Cournot oligopoly game, there are equilibria very different from a simple two-fold repetition of the static Cournot outcome. For example, it may be an equilibrium for the firms to collude perfectly in the first period and then play the static Cournot equilibrium in the second period.

[52] Nor can the oligopolists' customers make commitments or investments, such as building up their inventories when prices are low. In fact, the specification of repeated games requires that demand be intertemporally separable.

To highlight the difference between Nash and subgame perfect equilibria, and to introduce the issues involved in repeated oligopoly, I now characterize the set of symmetric Nash equilibria in this twice-repeated quantity game and compare it with the set of subgame perfect equilibria. I shall show that there is a range of Nash equilibria, only one of which is a subgame perfect equilibrium. All of the rest of the Nash equilibria are supported by incredible threats.

I begin with the Nash equilibria. Each firm's strategy consists of two parts: its first-period output, and its second-period output conditional on its rivals' first-period outputs. Let demand be given by $p(X)$, and assume for simplicity that the firms have constant and equal marginal costs of c. Let firm i's overall payoff be $\pi_{i1} + \delta\pi_{i2}$, where π_{it} is its payoff during period t and δ captures the relative importance of the second period.

First, notice that *any* equilibrium must involve static Cournot behavior in the second period; otherwise some firm would have an incentive to alter its second-period behavior. Call π_i^{C} firm i's (unique) static Cournot payoff. The range of Nash equilibria arises because of different *first-period* behavior in different equilibria.

Let $\pi(x) \equiv xp(nx) - cx$ be a single firm's profits if each of the n oligopolists selects output x. Call $\pi^{\mathrm{d}}(x) \equiv \max_y p(y + (n-1)x)y - cy$, the maximal profits that firm i can earn in a single period while deviating, given that all other firms are producing x. Now consider what output levels x can be supported as equilibrium choices during the first period. The best chance of supporting a particular x is to specify that deviations from producing x will be met by as severe a punishment as possible. The most severe punishment is to flood the market in the subsequent period (for example, by producing enough to drive price down to cost), thereby giving the deviating firm a second-period payoff of zero. Facing this (incredible) threat, a firm will comply with the specified equilibrium strategy of producing x during the initial period if and only if its equilibrium payoff, $\pi(x) + \delta\pi^{\mathrm{C}}$, is at least as large as its payoff from deviating, $\pi^{\mathrm{d}}(x)$. Any x for which the inequality

$$\pi(x) + \delta\pi^{\mathrm{C}} \geq \pi^{\mathrm{d}}(x) \tag{17}$$

can be supported as the first-period output in a symmetric Nash equilibrium.

Equation (17) typically will be satisfied for a range of outputs both greater and less than the Cournot level, x^{C}. This range is increasing in the discount factor, δ, and will include the fully collusive outcome if δ is large enough. For example, if $c = 0$ and $p(X) = 1 - X$, then $x^{\mathrm{C}} = 1/(n+1)$, $\pi^{\mathrm{C}} = 1/(1+n)^2$, $\pi(x) = (1-nx)x$, and $\pi^{\mathrm{d}}(x) = (1-(n-1)x^2)/2$. In particular, when $n = 2$ and $\delta = 1$, the Cournot output is $1/3$, whereas inequality (17) requires that x fall between $1/9$ and $5/9$.

Observe how the Nash equilibrium solution concept permits the incredible threat of flooding the market to be brandished in an indiscriminate fashion. A firm deviating from x during the initial period will be punished quite independently of the direction of the actual deviation from equilibrium or of any costs associated with executing the threat. For example, it is an equilibrium for the firms to produce so much in the first period as to drive price below cost, with each firm fearing that if it were to reduce its first-period output (an action that actually helps its rivals), the rivals will retaliate by flooding the market in the second period, eliminating the Cournot profits that otherwise would have been earned. Likewise, collusion during the first period can be achieved only by using the incredible threat of a price war during the second period if a firm deviates from collusion.

In fact, all of these Nash equilibria but one are supported by incredible threats. For example, in the collusive equilibrium described above, it is not credible that one's rival will flood the market if one defects during the initial period. Once the first period is past, it will not be in any firm's interest to carry out such a threat. The equilibrium concept of subgame perfection requires that every out-of-equilibrium strategy (threat *or* promise) be credible in the following sense: future strategies out of equilibrium (in our case, second-period behavior in the event of defection from first-period collusion) must *themselves* constitute an equilibrium in the future (sub)game.

Since the only equilibrium outcome in the second-period subgame of the twice-repeated Cournot game is the static Cournot outcome, *whatever* happened during the first period, any threat to behave differently in the future is not credible. Therefore, once we restrict our attention to credible threats, i.e. once we refine our equilibrium concept from Nash equilibrium to subgame perfect equilibrium, we find a unique equilibrium in the twice-repeated Cournot game (so long as the one-shot game itself has a unique equilibrium). Using the standard method of solving such a game backwards (just as one solves a finite optimal control problem backwards), it becomes clear that the unique equilibrium (henceforth, the modifier "subgame perfect" is implicit) is simply a two-fold repetition of the static Cournot outcome. Since no *credible* punishment for defectors is possible, it is not possible to support any first-period outcome other than the standard Cournot equilibrium. There are in fact no linkages in behavior between the two periods.

The preceding analysis points out the importance of the distinction between Nash equilibrium and subgame perfect equilibrium in dynamic games of all sorts, and in particular in the study of dynamic oligopoly. Although the perfect equilibrium solution concept is not without its own problems, it is now quite clear that for applications to oligopoly it is an essential refinement of Nash equilibrium that the serious student of oligopoly theory must thoroughly understand.

I have just shown that adding a second period to the static Cournot model, and refining our solution concept to subgame perfect equilibrium, simply implies a two-fold repetition of Cournot behavior. This surprising and counterintuitive result carries over to any finite number of repetitions of the Cournot game. Indeed, it applies to finite repetitions of any static oligopoly game.[53] Quite generally, the unique subgame perfect equilibrium of a finitely repeated game with a unique Nash equilibrium in the stage game is a simple repetition of the stage-game equilibrium.[54]

3.1.2. Alternative approaches to finitely repeated oligopolistic rivalry

The lack of any dynamics in finitely repeated oligopoly games might be considered an indictment on the perfect equilibrium solution concept that generates this result, or, indeed, on the entire game-theoretic approach being adopted. I remark here on two types of attacks suggesting that it is misleading to focus on perfect equilibria in finitely repeated oligopoly games.

Within the game-theoretic paradigm, it is clear that the triviality of equilibria in finitely repeated games is not robust with respect to either the structure of the game or the equilibrium concept. If we permit the length of the game, T, to be infinite (or at least *possibly* infinite), we suddenly get many perfect equilibria, some of which are very collusive, rather than just repetitions of the static equilibrium (see Subsection 3.2 below). Technically, there is a discontinuity in the equilibrium structure as $T \to \infty$. We also introduce genuine dynamics into the equilibrium behavior of the firms if we admit for some incomplete information, e.g. regarding the firms' costs or demand conditions (although this modification need not support tacit collusion – see Sections 4 and 5 below). And if we look for *ε-perfect equilibria* – i.e. configurations in which no firm can gain more than $\varepsilon > 0$ by deviating – we again can support very different behavior, even for ε close to 0.[55]

Another way of circumventing the "paradox" of the triviality of perfect equilibria in finitely repeated games is to throw out the very strong assumption

[53] This fact undermines earlier efforts to improve on the one-shot Cournot game by introducing reactions into finitely repeated quantity games. Friedman's (1968) reaction function equilibria are not subgame perfect.

[54] The qualification that there be a unique equilibrium in the one-shot game is essential. With multiple equilibria in the one-shot game, it is credible to "threaten" to revert to a less favorable equilibrium for the remaining periods in the event of defection. I regard these threats as an artifact of the definition of subgame perfect equilibrium, especially if the multiple equilibria can be Pareto ranked. (Why would the players not return to a Pareto-preferred equilibrium?) See the material below on renegotiation for a further development of this theme. See Benoit and Krishna (1985) for a thorough analysis of finitely repeated games with multiple equilibria, and Friedman (1985) for a similar study focused more explicitly on oligopoly.

[55] In fact, as Radner (1980) shows, as the length of the game becomes large, it is an ε-perfect equilibrium for the firms to collude for an arbitrarily long time.

that all firms have perfect computational abilities, act with perfect rationality and foresight, and know with certainty that the same is true of all of their competitors. At one level, this alternative approach is simply a recognition that the world is complex, and one cannot simply rely on the backwards induction analysis to conclude that repetitions of the static equilibrium strategies are optimal. Rosenthal (1981), for example, suggests that each firm might have some subjective beliefs regarding its rivals' future actions, not being confident that they will perform the flawless calculations underlying the perfect equilibrium.[56] Surely any businessperson finds the competitive environment far too complex to rely on the chain of calculations necessary in our perfect equilibrium calculation above.

This objection to formal game theory is borne out in the experimental evidence regarding finitely repeated games. The experimental evidence is strongest and clearest in the case of the finitely repeated prisoner's dilemma. Of course, this simple game is a special case of a repeated oligopoly game. Axelrod (1984) gives some of the most compelling evidence that cooperation *does* emerge in this game, even among very sophisticated players. He suggests an evolutionary approach to the problem rather than a game-theoretic approach. In particular, the "tit-for-tat" strategy of beginning with cooperation and then matching one's rival's previous move has extremely attractive survival properties in environments consisting of a variety of alternative rules.[57] "Tit-for-tat" does not thrive by besting its rivals, something it can never do, or by virtue of being the optimal response to *any* rival rule, but rather by finding a mix of cooperation and punishment.

3.2. Infinitely repeated oligopoly games: Supergames

It would appear that our attempt to use the theory of repeated games in order to focus purely on strategic interactions over time has completely failed to bring any such reactions into our analysis. Either we restrict attention to credible threats, in which case only the trivial repetitions of static equilibria are equilibria in finitely repeated games, or we do not, in which case we can support a wide range of behavior as Nash equilibria.

But a closer look at the result cited above, and the arguments supporting it, reveal a peculiar strength to the backwards induction or unraveling argument which narrowed down the equilibria to one. Threats of all sorts were not credible

[56] This approach has been pursued *within* the formal confines of game theory by examining games of incomplete information, as in, for example, Kreps and Wilson (1982b). But other ways of approaching this problem may prove even more fruitful.

[57] Survival is defined in an evolutionary sense: the population of players adopting a given rule or strategy grows or shrinks over time according to the performance of those players adopting the rule. For a much richer study of evolutionary processes in oligopolistic competition, see Nelson and Winter (1982).

in the final period, and hence in the penultimate period, and so on. The inability of the firms to tacitly collude in a credible fashion rests very heavily on the exogenously given terminal date at which rivalry ends. In reality, of course, competition continues indefinitely, or at least the firms cannot be sure just when it will end. What happens if one developes a theory of repeated rivalry without the artificial device of a known, finite end date?

This question has spawned a large literature on *infinitely* repeated oligopoly games. In fact, these games have exactly the opposite problem to the finitely repeated games when it comes to developing a useful, predictive theory of oligopoly: there is generally a plethora of perfect equilibria. This is especially true when there is no discounting, but remains the case when there is discounting if the per-period discount factor used to convert a stream of profits into a firm's payoff, δ, $0 < \delta < 1$, is close to unity. A high discount factor corresponds to short periods in a game that is likely to continue for a long time.[58]

Infinitely repeated games are fundamentally different from finitely repeated ones in that there is always the possibility of retaliation and punishment in the future. With an infinite horizon, the requirement that threats be credible turns out to be much less restrictive than it was in finite horizon games. Formally, a supergame consists of an infinite number of repetitions of some stage game. The repetitions take place at dates $t = 0, 1, \ldots,$ with player i's action at date t denoted by x_{it}, $i = 1, \ldots, n$. Denote by $\pi_{it} \equiv \pi_i(\boldsymbol{x}_t)$ player i's period-t payoff if the vector of actions $\boldsymbol{x}_t \equiv (x_{1t}, x_{2t}, \ldots, x_{nt})$ is played at date t. Player i's overall payoff in the game is given by[59]

$$\pi_i = \sum_{t=0}^{\infty} \delta^t \pi_{it}.$$

Note that the history of the game up to date t has no direct impact on either the payoffs or the feasible strategies from date t onwards. The game beginning at date t looks the same for all t, in the sense that the feasible strategies and the prospective payoffs that they induce are always the same. This stationarity reflects the absence of any investments by the firms (or any changes in the underlying competitive environment). History matters only because the firms

[58]Formally, δ may be thought of as the product of two terms: $\delta = \mu e^{-iT}$, where μ is the hazard rate for the competition continuing (i.e. the probability that the game continues after a given period, given that it has not previously ended), and e^{-iT} is the pure interest component of the discount factor, with period length T and interest rate i. In most papers, it is assumed that the competition surely continues indefinitely, i.e. $\mu = 1$, so that $\delta = e^{-iT}$. But it is worth noting that (with risk-neutral firms) these models are equally able to handle the case of an uncertain terminal date, so long as the hazard rate is constant.

[59]Many of the supergame results below generalize quite easily to the case of discount factors that vary, both across firms and with time. Friedman (1971), for example, allows the payoffs to be of the form $\pi_i = \sum \alpha_{it} \pi_{it}$.

remember what has happened in the past and condition their current actions on previous behavior. In this sense, any tacit collusion in supergames is of a "pure bootstrapping" variety; history matters because the players decide that it matters.

Supergames are well suited for the exploration of the efficacy of tacit collusion. Indeed, one of the primary questions addressed in the oligopolistic supergame literature is the following: Can the oligopolists, without any explicit collusion, support the profit-maximizing outcome (or any given Pareto optimal outcome) purely with credible threats to punish any defector who failed to cooperate with the proposed collusive arrangement?

3.2.1. *Tacit collusion with reversion to noncooperative behavior*

Friedman's (1971) important paper demonstrated that tacit collusion can indeed support any Pareto optimal outcome in an oligopolistic supergame if δ is close to unity, as it will be if the market participants can respond rapidly to each other. Friedman's equilibrium involves the following strategies: each firm produces its share of the collusive output in each period, so long as all others continue to do so. If, however, any firm produces more than its "quota" under this arrangement, this defection signals a collapse of a tacitly collusive arrangement, and each firm plays its (static) noncooperative strategy thereafter. In the case of quantity-setting supergame, i.e. repeated Cournot, each firm plays its static Cournot output following any deviation.

If these strategies indeed form a subgame perfect equilibrium, then tacit collusion works perfectly. We need to check that no firm would want to defect from the collusive scheme, and that the punishment strategies are themselves credible. Credibility requires that the punishment itself form a perfect equilibrium. A punishment involving repetition of an equilibrium in the stage game is credible, since it is always a perfect equilibrium to simply repeat any equilibrium in the stage game indefinitely. In quantity-setting supergames, this type of punishment is known as *Cournot reversion*.

To check the "no defection" condition for firm i, denote that firm's flow of profits under a candidate tacitly collusive equilibrium by π_i^*,[60] and its profits during the period in which it deviates from this collusive scheme by π_i^{d}. Call firm i's (flow) profits during the infinitely long *punishment phase* π_i^{p}; π_i^{p} may simply be the Cournot profit level, but we will consider more complex and severe punishments below. Firm i earns $\pi_i^*/(1-\delta)$ by cooperating. We have an equilibrium so long as this is no less than the deviant's profits, $\pi_i^{\mathrm{d}} + \delta\pi_i^{\mathrm{p}}/(1-\delta)$,

[60] Of special interest will be the case in which $\pi_i^* = \pi_i^{\mathrm{c}}$, firm i's profits when the firms collude fully to maximize their joint profits (if side payments are feasible) or to achieve a point on their Pareto frontier. But the calculations below apply to *any* candidate supergame equilibrium with so-called "grim" strategies calling for perpetual punishment of any defector.

for every firm i. Simple algebra indicates that this condition is equivalent to

$$\delta \geq \frac{\pi_i^{\mathrm{d}} - \pi_i^*}{\pi_i^{\mathrm{d}} - \pi_i^{\mathrm{p}}}, \quad i = 1, \ldots, n. \tag{18}$$

Under quite general conditions, each firm earns more during periods of tacit collusion than during the punishment phase, $\pi_i^* > \pi_i^{\mathrm{p}}$ for all i, and the profits that a defector earns are bounded, $\pi_i^{\mathrm{d}} < \infty$ for all i. If these conditions are satisfied, then equation (18) must be met for every firm i if δ is close enough to unity. In particular, any Pareto optimal set of outputs, or any set of outputs that maximizes joint profits, is supportable as a subgame perfect equilibrium. Basically, any short-run gains from defection, $\pi_i^{\mathrm{d}} - \pi_i^*$, are necessarily outweighed by the loss forever after of $\pi_i^* - \pi_i^{\mathrm{p}}$, i.e. by the collapse of the tacitly collusive arrangement. Equation (18) tells us that we need only mild punishments ($\pi_i^{\mathrm{p}} < \pi_i^*$) and bounded profits from defection ($\pi_i^{\mathrm{d}} < \infty$) if δ is close enough to unity.

This leads us to one of the most important conclusions with genuine policy implications that comes out of oligopoly theory. Whatever one believes about the various π's, it is clear that lower values of δ inhibit tacit collusion. If industry behavior permits each oligopolist to rapidly and surely observe rival defections, the scope for tacit collusion is great. Policies designed to make secret price cuts possible are valuable in undermining tacit collusion, or "conscious parallelism". And industry practices that inhibit secret price-cutting should be subject to close antitrust scrutiny.[61]

The feasibility of tacit collusion when detection lags are short is rather sobering for those who would conclude on theoretical grounds that oligopolistic behavior tends to be quite competitive. After all, supergame theory tells us that the fully collusive outcome is an equilibrium, quite independently of demand conditions or the number of oligopolists, so long as firms can rapidly detect and respond to "cheating" on the tacitly collusive scheme. The lower prices threatened as a price war never actually are charged. Worse yet from the point of view of industry performance, structural remedies would appear to hold out little hope of undermining tacit collusion if swift reversion to a noncooperative equilibrium is possible.[62] Nor need the entry of more firms improve industry behavior.[63]

[61]Again I refer the reader to Stigler's (1964) article for a discussion of the problems of policing a tacitly collusive scheme. See also the subsection below on trigger price strategies.

[62]But note the peculiar character of these supergame equilibria: firms believe that swift and perpetual punishment will follow if they defect, but defection and punishment never occur in equilibrium. Clearly, one should check the robustness of the results presented here to the presence of some noise that may cause price wars to actually break out in equilibrium. The literature on trigger price strategies undertakes this task.

[63]On the contrary, the incumbent firms may dissipate their collusive profits in deterring entry. Alternatively, entry, or nonprice competition, may dissipate the profits by raising industrywide average costs.

Although (18) is always met for δ close to unity, it is of some interest to explore the factors that tend to make tacit collusion successful. The simplest way to measure "success" is by finding the highest level of profit for firm i, π_i^* that can be supported as an equilibrium.[64] In terms of the reduced-form π's, equation (18) indicates that a higher value of π_i^* can be maintained in equilibrium if defection is less profitable (π_i^{d} smaller), if punishment is more severe (π_i^{p} smaller), or if detection and punishment are swifter (δ larger).

All of these findings are intuitive at this reduced-form level, but some of their implications in terms of the underlying structural variables are not. Generally, any underlying market condition that makes *very* competitive behavior possible and credible can, by lowering π_i^{p}, actually promote collusion. This is the "topsy-turvy" principle of supergame theory to which I alluded above.[65]

Here are two applications of the topsy-turvy principle. On the cost side, we usually think of flat marginal cost curves as leading to relatively competitive behavior. But the ability to rapidly expand production allows firms to punish a defector more harshly, and can promote collusion. Or, as Rotemberg and Saloner (1986b) point out, a quota on imports may make it impossible for a foreign firm to punish a domestic defector, and hence the quota may enhance competition. One must exercise extreme caution is using counterintuitive results of this sort that are based on the topsy-turvy principle. In the quota case, for example, the quota would have to be nonbinding in equilibrium for it to serve as an effective threat; in practice, quotas are binding and appear not to play the suggested pro-competitive role.

What does supergame theory tell us about the relationship between market structure and tacit collusion? In particular, how does the success of tacit collusion depend upon the number of firms, n? In a symmetric, quantity-setting supergame, an increase in n lowers each firm's share of the collusive profits, π_i^*, lowers its profits following Cournot reversion, π_i^{p}, and lowers its profits during defection, π_i^{d} [since its rivals together produce a fraction $(n-1)/n$ of the collusive output]. For the example of linear demand, $p = A - X$, and constant marginal costs $c < A$, direct substitutions into equation (18) tell us that tacit collusion with Cournot reversion supports the monopoly output level if and only if

$$\delta \geq \frac{(n+1)^2 - 4n}{(n+1)^2 - 16n^2/(n+1)^2}.$$

If we take $\delta = 0.99$ to approximate monthly detection lags and an annual interest

[64] In fact, much of the literature focuses on a narrower question of when tacit collusion works perfectly: "When is joint profit maximization sustainable as an equilibrium?"

[65] Avinash Dixit has noted the "Orwellian" character of the topsy-turvy principle, suggesting the theme "competition is collusion".

rate of 12 percent, then tacit collusion works perfectly so long as there are no more than 400 firms! This calculation is again discouraging for those who would believe on the basis of economic theory that oligopolies perform competitively. If one believes that oligopolists can somehow coordinate on a Pareto optimal supergame equilibrium, if one takes realistic values of the key parameters, and if one assumes that a firm can accurately observe its rivals' defections, then no firm would find the short-run gains from defection large enough to justify the long-run breakdown of tacit collusion.

Friedman's result that the collusive outcome can be sustained as a noncooperative equilibrium if δ is close to unity is actually a special case of a general result in supergames. Under quite general conditions, repetition of *any* individually rational outcome in the stage game can be supported as a supergame equilibrium with sufficiently little discounting.[66] The supporting strategies specify that any defecting firm be punished (by having the other firms minmax that firm), that any firm failing to participate in the punishment of another be punished, etc. See Fudenberg and Maskin (1986) for an excellent treatment of such "Folk theorems". The great multiplicity of supergame equilibria when δ is close to unity is a drawback to the entire supergame literature which I shall discuss at greater length below.

The presence of so many supergame equilibria when δ is close to unity suggests two approaches if one is to develop further the theory of oligopolistic supergames. The first is to examine the equilibrium structure of such games when δ is *not* close to unity, i.e. in circumstances where there are significant lags in reacting to rivals or the game is likely to end at a relatively early date. The second approach is to abandon the assumption of rapid and flawless observability of defections. I discuss these approaches in turn. Under either approach, there remain many supergame equilibria, so the natural question continues to be that of how collusive an outcome can be supported. When this question is answered, we can examine how its answer varies with structural parameters.

3.2.2. *Pareto optimal equilibria in quantity-setting supergames*

Abreu (1986) has recently provided us with an important advance in our understanding of the structure of pure strategy equilibria for oligopolistic supergames with discounting. He seeks to identify (p. 192) "the *maximal* degree of collusion sustainable by credible threats for arbitrary values of the discount factor". As discussed above, to achieve maximal credible collusion, one must design the most severe credible punishments for defectors. To make a punish-

[66]An outcome is individually rational for firm i if it gives that firm a payoff no lower than the one that firm i could guarantee itself against any play by the other $n - 1$ firms. In other words, firm i must earn at least its *minmax* level of profits.

ment regime credible, it is necessary to punish a player who fails to participate in the punishment of another player. Using this idea, Abreu has been able to characterize quite generally the optimal punishment strategies, which he calls optimal penal codes, and hence the most collusive perfect equilibria. As we shall see, the Cournot reversion discussed above is *not* in general an optimal punishment, and hence cannot support the most collusive equilibria.[67]

I begin by reporting Abreu's (1988) general results about supergames with discounting. He shows that the key to characterizing the Pareto optimal equilibrium points is to construct the perfect equilibrium yielding player i the lowest payoff, for $i = 1, \ldots, n$. Then one supports the most collusive outcomes using equilibrium strategies that call for a reversion to player i's least-preferred perfect equilibrium if that player defects from the strategies specified in the initial equilibrium. And, importantly, if player j does not participate in the punishment of player i (i may equal j here), the strategies specify that player j will then be punished via reversion to *his* least-preferred perfect equilibrium. Effectively, Abreu has shown that one may restrict attention to strategies in which history matters only through the identity of the most recent defector (where failure to properly punish a previous defector itself counts as defection).

In this way, Abreu manages to simplify the problem to one of identifying the n least-preferred perfect equilibria, for players $i = 1, \ldots, n$. When these equilibria are identified, call w_i the payoff to player i in his worst perfect equilibrium. Then an outcome in which player i earns $\pi_i{}^*$ each period, but could defect and earn $\pi_i{}^{\mathrm{d}}$ during a single period, is an equilibrium if and only if $\pi_i{}^*/(1 - \delta)$ is no less than $\pi^{\mathrm{d}} + \delta w_i$ for all $i = 1, \ldots, n$.

In symmetric games, the n punishment equilibria are simply permutations of one another, distinguished on the basis of which player is being punished. Punishment is supported by the threat to impose this single worst equilibrium path on any firm defecting from the prevailing punishment. So, in symmetric games, the key to identifying most-collusive outcomes is to find the single equilibrium path yielding player one, say, the lowest payoff.

In applying his general findings to symmetric oligopolistic games, Abreu's strongest results come when he restricts attention to symmetric punishments, i.e. punishments which specify that all firms act identically. As he points out, (1986, p. 198), "the restriction to symmetric paths is neither natural nor in principle innocuous". Despite the symmetry of the overall game, the presence of a defector destroys the symmetry when it is necessary to mete out punishments. Symmetric punishments are nonetheless a natural generalization of Cournot reversion, since Cournot reversion is itself symmetric.

[67]Of course, for δ close to unity, Cournot reversion is optimal. The point is that more sophisticated punishments expand the range of discount factors over which full collusion can be supported, and support more collusive outcomes when δ is smaller yet.

Within the class of symmetric punishments, Abreu proves that the optimal punishment has a simple, two-phase structure: immediately following the defection, each firm participates in a "price war" by producing a higher output than previously; but immediately thereafter all firms return to their optimal, tacitly-collusive output levels. It is striking that, when optimally punishing a defector, the industry returns after only a single period to the most collusive sustainable configuration. Abreu describes these types of punishments as offering a stick and a carrot; apparently, the carrot (returning to collusion) is necessary to make the stick (the one-period price war) both credible and as menacing as possible.

Abreu's intuitive argument establishing the optimality of two-phase symmetric punishments is simple and instructive. Consider any arbitrary (but credible) punishment path Y calling for per-firm outputs of y_t at dates $t = 1, \dots$. Call the most collusive supportable per-firm output level x^*. Consider replacing Y by a two-stage path X having per-firm output x^{p} at $t = 1$, and x^* thereafter. The path X consists of a price war followed by a return to tacit collusion. Producing x^* forever yields each firm a payoff at least as high as does $\{y_2, y_3, \dots\}$, so we can select $x^{\mathrm{p}} > y_1$ to make the overall per-firm payoffs under X equal to those under Y. Then X can be used as punishment path in place of Y so long as it too is a perfect equilibrium. Since playing x^* indefinitely is by definition supportable, the only thing to check is that no firm would want to defect during the price war. But, since $x^{\mathrm{p}} > y_1$, defecting from X during the first period is less attractive than defecting from Y, which itself was unprofitable (since Y forms a perfect equilibrium). And by construction the two paths yield the same payoff from compliance. Therefore, each firm finds it optimal to participate in the price war by producing x^{p}.

Note that each firm must earn lower profits during the price war than it would in the static Cournot equilibrium, since its overall payoff under the optimal punishment is lower than under Cournot reversion, despite its higher profits during the second phase of the punishment. In fact, each firm may easily lose money during the price war, especially if δ is moderately large. All that is required for individual rationality is that no firm lose money prospectively at any time, looking ahead to both the price war and the subsequent return to collusion.

Not only do the optimal symmetric punishments take on this simple two-stage form. They are about as easy to calculate as one could expect, given the complexity of the problem. Let $\pi(x) \equiv p(nx)x - cx$ be the per-firm profit when each firm produces x, and call $\pi^{\mathrm{d}}(x)$ the maximal profits that firm i can earn in a single period while deviating, given that all other firms are producing x.[68] Suppose that the monopoly output is not supportable as an equilibrium. Then the best collusive output, x^*, and the price-war output x^{p}, are defined by the two

[68] Formally, $\pi^{\mathrm{d}}(x) = \max_y p(y + (n-1)x)y - cy$.

equations:

$$\pi^{\mathrm{d}}(x^*) - \pi(x^*) = \delta(\pi(x^*) - \pi(x^{\mathrm{p}})) \tag{19}$$

and

$$\pi^{\mathrm{d}}(x^p) - \pi(x^{\mathrm{p}}) = \delta(\pi(x^*) - \pi(x^{\mathrm{p}})). \tag{20}$$

The first of these two equations is our familiar no-defection condition. In this case it requires that each firm just be indifferent between tacitly colluding and defecting.[69] The left-hand side of (19) gives the benefit of defection, and the right-hand side the cost, namely the lost profits due to the single-period price war. The second equation requires that each firm be willing to go along with the punishment, realizing that failure to do so would simply extend the price war for another period. The left-hand side of (20) is the gain from defection during the price war, and the right-hand side is the loss tomorrow from having a price war rather than collusion at that time. As Abreu emphasizes, one needs only to look ahead a single period in order to evaluate the attractiveness of any (possibly long and complicated) deviation.

One reason why Abreu's work is important to our understanding of tacit collusion is that he proves quite generally that the optimal punishment strategies are more severe than Cournot reversion. In fact, Abreu shows the following: so long as Cournot reversion supports an outcome that is more collusive than Cournot itself, Cournot reversion cannot be the most severe credible symmetric punishment. Quite generally, a firm can credibly be made worse off than it would be by simple repetition of the Cournot equilibrium. Therefore, unless Cournot reversion itself supports the monopoly outcome, a greater degree of collusion can be sustained using Abreu's stick and carrot strategies than simply through threats to revert to Cournot behavior.

Abreu also demonstrates conditions under which symmetric punishments (and, hence, two-phase punishments) are fully optimal. This occurs if and only if they can support continuation (punishment) equilibria in which the firms earn zero profits – clearly the lowest possible. For large enough values of δ, this is always the case. Otherwise, however, Abreu shows that asymmetric punishments do strictly better. An interesting feature of asymmetric punishments is that the firms meting out the punishment produce strictly higher output, during the initial period of punishment, than does the firm being punished. It also is generally true that the firm being punished "cooperates" with its punishment, in the sense that

[69] If the monopoly output *can* be supported, then we must set x^* in equation (19) equal to x^{c}, each firm's share of the fully collusive, monopoly output, and replace the equality in equation (19) by an inequality. With these changes, equation (19) reads $\pi^{\mathrm{d}}(x^{\mathrm{c}}) - \pi(x^{\mathrm{c}}) \leq \delta(\pi(x^{\mathrm{c}}) - \pi(x^{\mathrm{p}}))$. See Abreu (1986, p. 203).

it does not select an output along its static Cournot reaction schedule. Abreu concludes with some statements and some conjectures about the nonstationarity of optimal asymmetric punishments.

3.2.3. Price-setting supergames

Much of the literature on oligopolistic supergames follows the discussion above in looking at quantity-setting supergames. This is natural in view of the advantages of the Cournot model over the Bertrand model for homogeneous goods. Repeated Bertrand games, i.e. price-setting supergames, are however an obvious alternative approach. The general theorems cited above regarding supergame equilibria when $\delta \to 1$ tell us that collusion is equally well supportable in price games as in quantity games when responses are rapid.

Generally, the fact that the static Bertrand equilibrium is so competitive means that punishments in pricing games are severe. This can make it *easier* to support collusion in a price-setting supergame than in the related quantity-setting supergame.[70] Again we have an application of the topsy-turvy principle of supergames. Reversion to the static Bertrand equilibrium is a more severe punishment than is Cournot reversion, and may support more collusive behavior. In fact, the distinction between noncooperative reversion and optimal punishments is greatly muted in price-setting supergames.

In the central case of homogeneous goods and constant marginal costs, Bertrand reversion gives the firms zero profits, and therefore must be the optimal punishment. Setting $\pi^{\mathrm{p}} = 0$, our fundamental no-defection condition, equation (18) becomes:

$$\pi_i^{\mathrm{d}} \leq \frac{\pi_i^*}{1-\delta}, \tag{21}$$

where the left-hand side of (21) is the total payoff from defection and the right-hand side is the payoff from cooperation. In the symmetric case, we can say much more about the relationship between π_i^{d} and π_i^*. Optimal defection involes a slight reduction in price to capture the entire market. This strategy garners profits of $\pi_i^{\mathrm{d}} = n\pi_i^*$ for firm i, since it expands its sales by a factor of n. Making this substitution, we see that equation (21) simplifies to

$$n(1-\delta) \leq 1. \tag{22}$$

The efficacy of tacit collusion depends not at all on the price chosen, since π_i^*

[70] But this effect must be balanced against the fact that deviation is generally more profitable under Bertrand than under Cournot behavior.

has disappeared, but only on the number of firms and the speed of retaliation.[71] Therefore, the oligopolists will support the monopoly price if they can tacitly collude at all. Again, for plausible parameters, tacit collusion would appear to work well; for $\delta = 0.99$, equation (22) is satisfied for any $n < 100$.

Brock and Scheinkman (1985) analyze Bertrand reversion without the special assumption of constant marginal costs. Instead, they assume that the firms produce subject to capacity constraints. In this case, Bertrand reversion means playing the mixed strategies that constitute the static equilibrium (since we know from Subsection 2.2 above that there is generally no pure strategy Bertrand equilibrium with capacity constraints). Even in this case, however, Bertrand reversion is the worst that could happen to a given firm (i.e. each firm earns only its minmax payoff) since each firm's expected profits in the mixed strategy equilibrium equal those that it would earn if each rival were producing at capacity and the given firm were optimizing against the resulting residual demand.[72] In this context, Brock and Scheinkman ask how the degree of sustainable collusion relates to the number of firms. Without capacity constraints, we can see from equation (22) that an increase in n reduces the degree of sustainable collusion, because each firm finds defection more tempting when its own market share is low.[73] But this monotonicity result may fail in the presence of capacity constraints; with limited capacities, a larger number of firms lowers the punishment profits, π_i^{p}, and hence may support greater collusion. Again the topsy-turvy principle rears its head. And again one must be careful in applying the topsy-turvy result. In this case, an increase in the number of firms could promote tacit collusion only in an oligopolistic industry with perpetually unused capacity.

Rotemberg and Saloner (1986a) add an interesting wrinkle to these models of price-setting supergames. They study tacit collusion in the presence of observable but temporary shifts in industry demand. Their goal is to explain oligopoly behavior over the course of the business cycle. Each period the oligopolists observe the current state of demand, as summarized by a realization of the random variable $\tilde{\theta}$, where demand increases with θ. Crucially, the random shocks to demand exhibit no serial correlation. With this assumption, today's demand conditions convey no information about future demand, so the *future* always looks the same, although the current-period payoffs depend upon the current state of demand. Call $\bar{\pi}_i^*$ firm i's expected flow profits under tacit collusion, where the expectation is taken over $\tilde{\theta}$. Denote by $\pi_i^*(\theta)$ and $\pi_i^{\mathrm{d}}(\theta)$ firm i's current profits from cooperating and defecting, respectively, and when the state

[71]The reason for this strong result is that higher prices raise *both* the profits from defection and the profits from cooperation. In the special case of constant marginal costs and unlimited capacities, these effects just cancel out to give equation (22).

[72]See Brock and Scheinkman (1985, p. 373).

[73]At any price, an increase in n raises π_i^{d} but does not alter π_i^{p}, which is uniformly zero.

of demand is θ. These profits depend of course upon the price specified by the tacitly collusive scheme when the state of demand is θ.

With constant marginal costs, Bertrand reversion generates zero profits, $\pi_i^{\mathrm{p}} = 0$, and defecting when demand is in state θ gives firm i an overall payoff of $\pi_i^{\mathrm{d}}(\theta)$. In contrast, cooperating yields a payoff of $\pi_i^*(\theta) + \delta\bar{\pi}_i^*/(1-\delta)$. Comparing these two payoffs, the no-defection condition, equation (18), becomes:

$$\pi_i^{\mathrm{d}}(\theta) - \pi_i^*(\theta) \leq \frac{\delta}{1-\delta}\bar{\pi}_i^*. \tag{23}$$

As we saw above, the gains from defection on the left-hand side of equation (23) are $(n-1)\pi_i^*(\theta)$, since cooperation gives a $1/n$ share of the market and profits of $\pi_i^*(\theta)$, while defection allows a firm to capture the entire market, earning profits of $n\pi_i^*(\theta)$. Making this substitution, equation (23) becomes simply:

$$(n-1)\pi_i^*(\theta) \leq \frac{\delta}{1-\delta}\bar{\pi}_i^*. \tag{24}$$

Equation (24) can be rewritten as

$$n(1-\delta) \leq 1 - \delta\frac{\pi_i^*(\theta) - \bar{\pi}_i^*}{\pi_i^*(\theta)}$$

to highlight the comparison with equation (22). Unlike equation (22), the π_i^* terms do not cancel in equation (24).

Now Rotemberg and Saloner's main point is that, for a given price, the left-hand side of equation (24), i.e. the gain from defection, increases with θ. This implies that tacit collusion becomes more difficult in high-demand states, in the sense that prices must be lowered when θ is large in order to sustain collusion. They conclude that the oligopolists are able to support less collusion during high-demand states than low-demand ones in order to prevent defection.[74] Note that the title's reference to "price wars during booms" is somewhat misleading, since prices are generally higher during business cycle peaks than troughs, and since price wars never occur in equilibrium. During booms, the oligopolists collude somewhat less in order to prevent the collapse of the implicit cartel.

Rotemberg and Saloner go on to discuss the implications of this finding for the behavior of oligopolies over the course of the business cycle. To the extent that

[74] This result does not carry over in general to the case of nonconstant costs, since both the optimal strategy when defecting and Bertrand reversion are quite different in that case. Nor does the incentive to defect necessarily increase with the state of demand for quantity-setting supergames. Rotemberg and Saloner do show, however, that their main point remains valid for both price and quantity competition if demand and marginal costs are linear.

oligopolists engage in less collusion during periods of high demand than during slack periods, their behavior accentuates the exogenously given shocks to demand. I find these results suggestive, but they must be treated with caution for several reasons. First, their predictions are at odds with the bulk of the empirical evidence on business cycles and oligopoly behavior. For example, Domowitz, Hubbard and Petersen (1986) report a positive sensitivity of price–cost margins to demand conditions that is most pronounced in highly concentrated industries. Scherer (1980) also cites evidence that undermines the "price wars during booms" theory.[75] Second, as Rotemberg and Saloner point out, capacity constraints are likely to be binding during periods of high demand, in which case defection is well-nigh impossible, and price-cutting is hardly necessary at these times in order to prevent defection. Third, Rotemberg and Saloner's assumption that demand shocks display no serial correlation seems inappropriate in a discussion of business cycles. The more natural assumption of white noise innovations in demand would likely give very different results, as (expected) future variables would move along with current conditions. Finally, this paper is subject to the general criticisms of supergames that I enumerate below.

3.2.4. *Oligopolistic supergames with imperfect monitoring*

So far, I have assumed that any deviation from the tacitly collusive scheme by firm i is immediately and accurately observed by all of that firm's rivals (although there may be a lag before they can respond). This assumption is clearly both crucial to the models above and unlikely to be met in practice. Indeed, efforts by firms to facilitate the exchange of pricing information via trade associations indicate that such information is not inevitably or typically available and that oligopolists value such information. Quite appropriately, antitrust law casts a skeptical eye on the exchange of customer-specific pricing information among oligopolists.

The question thus arises: *What happens to tacit collusion if oligopolists cannot easily observe rivals' price-cutting or production levels?*[76] In order formally to study this question, one needs a model of repeated rivalry in which the firms cannot perfectly observe, or infer, their rivals' actions.[77] The literature on *trigger*

[75] But Rotemberg and Saloner present some evidence of their own that supports their theory.

[76] A related question, "What happens to tacit collusion if oligopolists cannot easily observe each other's costs?", becomes relevant when firms' costs vary over time and these variations are not perfectly correlated across firms. As with unobservable production levels, unobservable costs pose a barrier to effective tacit collusion, since the firms as a whole would like to shift production towards the lower-cost firms at any given date, but each firm has an incentive to claim that its costs are low in order to have its allowable output increased. For the effect of exchanging cost information in a static model, see Shapiro (1986).

[77] There is a close relationship between oligopolistic supergames with imperfect monitoring and repeated moral hazard situations where each agent can temporarily gain by reducing effort, but may be punished if this shirking is observed by the other agents.

price strategies explicitly introduces informational imperfections that make it difficult for a firm to determine with certainty just when, or by how much, a rival has exceeded its collusive output level. This approach is the most direct descendent of Stigler's (1964) work on oligopoly theory.

Within the context of a homogeneous-goods, quantity-setting model, a natural way to incorporate imperfect monitoring of rivals is to assume that a firm can observe the market price in period t, p_t (which it receives for its products), but not the production levels of its rivals, x_{jt}, $j \neq i$. But with known demand, the firm could use the market price to infer the aggregate production of its rivals, using the identity $X_{-i} \equiv D(p) - x_i$, where $D(\cdot)$ is the direct demand function. And perfect information about rivals' production is exactly the assumption that we are trying to relax. Therefore, a coherent theory of imperfect cartel monitoring must incorporate demand (or cost) uncertainty. Demand shocks are then confounded with rivals' defections. With this structure, we can restate the cartel monitoring problem as the following question: *What happens in supergames if each oligopolist cannot distinguish with certainty between downturns in demand and expansion by a rival?*

Consider a quantity-setting supergame with stochastic demand. The simplest way to introduce demand uncertainty is to assume that demand in period t is of the form $p_t = \theta_t p(X)$, where θ_t is the realization of the demand shock at period t. To keep matters simple, assume that the θ_t's are drawn independently according to the cumulative distribution function $F(\theta)$. Without loss of generality we can take the expectation of θ to be unity. When one of the oligopolists observes a low price during period t, it cannot in general tell whether the low price is a consequence of a low realization of θ or of extra output by a rival. Importantly, we must assume firm i cannot observe its rivals' outputs either contemporaneously or subsequently.[78]

Green and Porter (1984) and Porter (1983a) have explored a class of supergame strategies known as *trigger price strategies* using this model. A trigger price strategy is a particular way of coordinating tacit collusion. Each firm produces its tacitly collusive or "cooperative" output, x^* until price falls below the trigger price, $\tilde{p}$, during some period. Any price below $\tilde{p}$ initiates a punishment phase or price war consisting of Cournot reversion for $T - 1$ periods. After T periods, the firms return to their original strategies, again cooperating at x^* until price falls below the trigger. Although trigger price strategies are not in general the optimal way to police tacit collusion (see below), they are relatively simple and have an element of realism to them. A trigger price scheme is characterized by three parameters: $\tilde{p}$, x^*, and T.

[78] I have already discussed what happens when firms can observe defections with a lag; this corresponds to a low value of δ.

Consider now the problem of selecting the trigger price parameters in order to generate the highest possible profits per firm.[79] In equilibrium, no firm will choose to defect from the tacitly collusive output x^*, but occasionally an unfavorable demand shock will drive prices below the trigger level and cause the firms to initiate a price war (Cournot reversion). The basic tradeoff is therefore the following: a low trigger price reduces the per-period probability of initiating a price war, but requires longer punishments or lower collusive profits in order to deter firms from cheating on the arrangement.

Denote by $\pi^* \equiv p(nx^*)x^* - C(x^*)$ a firm's expected profits during a period in which all firms produce x^*, and denote each firm's expected profits during Cournot reversion by π^{P}, in accordance with my earlier notation. Instead of working directly with the trigger price, $\tilde{p}$, it is useful to work with the induced (per-period) probability, $\tilde{\psi}$, of reverting to Cournot behavior. Cournot reversion occurs if and only if $p_t < \tilde{p}$. Given that in equilibrium each firm follows the tacitly collusive strategy of producing x^*, $p_t < \tilde{p}$ if and only if $\theta_t p(nx^*) < \tilde{p}$, which occurs with probability

$$\tilde{\psi} \equiv F\left(\frac{\tilde{p}}{p(nx^*)}\right). \tag{25}$$

Given $\tilde{\psi}$, x^*, and T, the overall payoff to each firm is given implicitly by[80]

$$V = \pi^* + (1 - \tilde{\psi})\delta V + \tilde{\psi}\left(\pi^P(\delta + \cdots + \delta^{T-1}) + \delta^T V\right). \tag{26}$$

The expected payoff V comprises the following terms: (i) the current expected payoff from cooperation, (ii) the payoff V beginning tomorrow, discounted by the probability, $1 - \tilde{\psi}$, that a price war will *not* be initiated during the current period, and (iii) the payoff upon beginning a price war, multiplied by the probability of that event, $\tilde{\psi}$. This last term is in turn made up of a stream of Cournot (punishment) payoffs for $T - 1$ periods followed by a return to tacit collusion. Solving equation (26) for V we have:

$$V = \frac{1}{1-\delta}\,\frac{\pi^*(1-\delta) + \pi^{\mathrm{P}}\tilde{\psi}(\delta - \delta^T)}{(1-\delta) + \tilde{\psi}(\delta - \delta^T)}. \tag{27}$$

Observe that $V(1 - \delta)$, the expected per-period payoff, is a weighted average of

[79]As we did above in the case of supergames with perfect monitoring, we are presuming that the firms can somehow coordinate and communicate at time zero to select a Pareto optimal sugergame equilibrium.

[80]This formula applies at the beginning of the game and also at any point when the firms are not engaged in a price war.

π^* and π^{p}, where the weights depend upon δ, $\tilde{\psi}$, and T. Equation (27) can be manipulated to yield:

$$V = \frac{\pi^{\mathrm{p}}}{1-\delta} + \frac{\pi^* - \pi^{\mathrm{p}}}{(1-\delta) + \tilde{\psi}(\delta - \delta^T)}, \tag{28}$$

which decomposes V into the returns from static, noncooperative behavior, $\pi^{\mathrm{p}}/(1-\delta)$, plus the single-period gains from colluding, appropriately discounted.

In order for a triplet $\{x^*, \tilde{\psi}, T\}$ to constitute a trigger price equilibrium, it must not be possible for any firm to raise its expected profits by producing more than x^*. In other words, we must have:

$$\left.\frac{\partial \pi_i}{\partial x_i}\right|_{x^*} \leq \delta \frac{\mathrm{d}\psi}{\mathrm{d}x_i}\left(\frac{\pi^* - \pi^{\mathrm{p}}}{1-\delta+\psi\delta}\right), \tag{29}$$

where $\psi \equiv F(\tilde{p}/p(x_i + (n-1)x^*))$ is the probability of a price war breaking out, so the $\mathrm{d}\psi/\mathrm{d}x_i$ term on the right-hand side of (29) measures the increased probability of a price war as x_i increases. The final term on the right-hand side of (29) is the cost to firm i of initiating a price war.

The (tacit) cartel management problem can now be stated as maximizing V in equation (28) subject to the inequality constraint in (29).[81] Equation (29) can be used to determine the smallest sustainable output x^* as a function of ψ and T, and then V can be maximized with respect to these variables.

In some ways, the most significant contribution of the papers by Green and Porter is in formalizing the notion of tacit collusion with imperfect price information. Some of their main "results" are really built into their definition of a trigger price equilibrium: (a) there are no defections in equilibrium, but price wars occur during periods of weak demand; the firms recognize that low prices are in fact due to unfavorable demand shocks, but carry out the specified punishments because they realize that failing to do so would cause the tacitly collusive scheme to collapse; (b) there are alternating phases of relatively collusive behavior and Cournot behavior – the firms begin by cooperating but inevitably experience price wars; (c) the collusive phases have random lengths; and (d) there is a single punishment – Cournot reversion for $T-1$ periods – for all crimes.

Green and Porter also provide some results characterizing the optimal trigger price strategies. First, Porter (1983a) shows that it is optimal for the firms to

[81] Of course, (29) is only a local condition, so one must really check that global deviations in x_i also are unprofitable.

produce in excess of the monopoly output in order to ease the problem of cartel enforcement; $x^* > x^c$ so long as $\delta < 1$.[82] Essentially, a small reduction in output below the monopoly level has no first-order effect on π^*, but it does reduce the gains from defection. Second, Green and Porter show that the optimal length of the punishment phase may be infinite, i.e. grim strategies may be optimal even in an uncertain environment. In fact, the optimal length of Cournot reversion is generally infinite [see Abreu, Pearce and Stacchetti (1986)]. Finally, Porter (1983b) presents some empirical support for the use of trigger price strategies by looking at the railroad industry in the 1880s.

Abreu, Pearce and Stacchetti (1986) provide a significant generalization of the Green and Porter analysis. They look for the fully optimal tacitly collusive equilibria with imperfect monitoring. They do not restrict attention to trigger strategies. In particular, they permit the punishment to depend upon the crime (the actual price observed), they allow for the possibility that prices prior to $t-1$ affect the behavior at time t, and they do not assume Cournot reversion as the mode of behavior during the punishment phase.

Abreu, Pearce and Stacchetti are able to derive some very powerful results characterizing the equilibria that are optimal among all pure-strategy symmetric sequential equilibria.[83] They show that there are only two (per-firm) output levels ever produced under the optimal scheme. One is the first-period output in the *best* symmetric sequential equilibrium, and the other is the first-period output in the *worst* symmetric sequential equilibrium. Each firm chooses between these two output levels on the basis of two factors: the output that was specified in the previous period, and the price that prevailed at that time. So, the firms need not look back further than one period and need not tailor the punishment to the price observed, even when such strategies are permitted. These results imply that the sequence of production levels is a Markov chain with only two states. The state is either cooperative or punitive. There is a transition rule from one state (output level) to another, depending upon the price observed.

Most recently, Abreu, Milgrom and Pearce (1987) have explored equilibria in the repeated prisoner's dilemma game using with a rather different (imperfect) monitoring technology. Although this work is at an early stage, it seems likely to have implications for oligopolistic supergames more generally. Their specification is designed to separate two aspects of repeated games with imperfect monitoring that typically are intertwined: (1) the frequency with which moves are made, and

[82] When $\delta = 1$, tacit collusion always works perfectly. See Radner (1986). Each firm can check statistically for defections by its rivals, reverting to punishment only when it becomes virtually certain that defection has occurred. This cautious punishment behavior could not deter cheating in the discounted game, but in the undiscounted game the magnitude of the punishment is not reduced merely because it is delayed.

[83] The restriction to symmetric punishments is less objectionable here than in Abreu's earlier work since no firm can determine which rival has defected.

(2) the lag with which firms receive information about their rivals' actions. In supergames with perfect monitoring, frequent opportunities to move necessarily go hand in hand with rapid observations of rival actions. In these games, we know that increasing the frequency of moves and observations, i.e. reducing the lag between moves, which is equivalent to raising δ, helps support collusion.[84] In the papers on imperfect monitoring discussed above, frequent moves go along with rapid, if imperfect, observations of defections. Again, increasing the frequency of moves and observations, as modeled by an increase in δ, helps support collusion.

Abreu, Milgrom and Pearce disentangle the timing of moves and the arrival of information by abandoning the assumption that even noisy industry aggregates are immediately observable to the oligopolists. Under their monitoring technology, the players may receive a public signal that carries information about whether anyone is defecting. In one version of the model, the signal is bad news: it is more likely to arrive if at least one firm is currently defecting than if none is.[85] In another version of the model, the signal is good news.

With this monitoring technology, varying the frequency of moves, t, is quite different from varying the rate at which the players discount the future. In particular, the limit as $t \to 0$ does not correspond to the limit as $\delta \to 1$: quick is not the same as patient. Although collusion is always possible in the limit as $\delta \to 1$, Abreu, Milgrom and Pearce's main result is that the best symmetric equilibrium payoff may be *increasing* in t for t close to zero. In other words, reducing the frequency of moves may make collusion more effective. The reason appears to be the following: although quicker moves increase the possibilities for punishment, they also increase the scope for defection. In the "bad news" model, this unexpected result is likely to arise near $t = 0$ if the players are patient or if the signal is much more likely to arrive during defection than during cooperation. In the "good news" model, only the static equilibrium (defection for all players) can be supported in the limit as $t \to 0$.

Abreu, Milgrom and Pearce's work is useful in focusing attention on the distinction between the timing of moves and the timing of information flows. Rather different technological factors determine these two aspects of oligopolistic interactions. Although the speedy arrival of accurate information about rival activities does increase the efficacy of collusion, it is not in general true that collusion is supported by an increase in the frequency with which firms can move.[86]

[84]And both of these changes are equivalent to a change in the firms' preferences that increase their patience.

[85]I.e. the Poisson process generating the signal has a higher arrival rate when not all firms are cooperating than when they are.

[86]Moreover, additional work by Abreu, Milgrom and Pearce suggests that increasing the lag with which imperfect information is observed by the players may enhance rather than undermine collusion.

The literature on repeated oligopoly games with demand uncertainty is a valuable extension of the basic supergame theory, and one that is likely to grow along with the related literature on repeated moral hazard. The key contribution to date of these theories is that they actually predict the occurrence of price wars in equilibrium. In contrast, in the previous supergame theories, while the credibility and size of punishments was critical, price wars never actually occurred.

3.2.5. Critique of supergames as a basis for oligopoly theory

Although the literature on quantity-setting supergames represents an enormous step beyond simple, static Cournot oligopoly, this literature suffers from several theoretical problems that limit its applicability and predictive power, even in industries where the environment might reasonably be taken as stationary. There are two major problems associated with oligopolistic supergames, each relating to the proper interpretation of supergame equilibria.

First, there is the difficulty of selecting among the vast multiplicity of equilibria. The huge number of supergame equilibria must be considered a major liability of this whole theoretical development. Certainly, game theory does not *predict* the collusive outcome; it simply indicates that such an outcome is supportable as a noncooperative equilibrium. The literature, and my survey of it, focuses on the set of Pareto optimal points within the larger set of equilibria, but in general there are still many Pareto optimal equilibrium points. Friedman (1971) developed the concept of *balanced temptation equilibrium* to select a point within the set of Pareto optimal equilibria.[87] but the balanced temptation criterion is itself ad hoc, and any less-collusive outcome also can be sustained using the same punishment strategies, since the gains from defection are smaller if the outcome is less collusive. Nor has any simple restriction on the firms' strategies yet been proposed that narrows down the equilibrium set in an instructive way.[88] So we have done no more than identified a large set of perfect equilibrium payoff vectors, and observed that the fully collusive point often is in this set.

Is there any reason to believe that the participants can focus on one of the Pareto preferred points in this equilibrium set? At the least, we have a bargaining problem among the firms, as each fights for a point on the Pareto frontier most favorable to itself. And bargaining theory allows for the possibility of an inefficient outcome. More generally, we have no compelling reason to rule out

[87] In my notation, a balanced temptation equilibrium is a Pareto optimal equilibrium at which the right-hand side of equation (18) is independent of i.

[88] For example, Stanford (1986) looks at "reaction function" strategies in a duopolistic supergame. These strategies require a firm to condition its period-t action solely on its rival's action during period $t-1$. Stanford shows that with this restriction the only subgame perfect equilibrium is simple repetition of the stage-game equilibrium.

Pareto dominated equilibrium points in games with multiple equilibria. While doing so has the advantage of narrowing down the set of equilibria that one must consider, it smacks of a cooperative behavioral axiom. And the firms may have difficulty adhering to the strategies necessary to support the most-collusive equilibria, at least in the case of complex, asymmetric punishments.

A second objection to supergame equilibria is that the punishments specified are *not* credible, even though we have restricted our attention to subgame perfect equilibria. This criticism goes back to the underlying interpretation of Nash equilibrium that was used to restrict attention to the Pareto optimal equilibria themselves. According to this interpretation of Nash equilibrium (which applies especially in the presence of multiple supergame equilibria), the participants communicate fully at time zero, specifying their behavior for the rest of the game. Then they go their separate ways, and implement what they have discussed. If the bargaining process does not fail at time zero, they should select a point on the Pareto frontier to enact. And so long as the specified strategies form a perfect equilibrium, it will not be in the interest of any party to deviate. The Nash equilibrium is a self-enforcing agreement, possibly a complex one calling for severe punishments for defectors.

Under this view of a Nash equilibrium as a self-enforcing agreement, however, it is not clear why the firms *jointly* would actually implement the severe punishments specified by their initial agreement. In other words, is it credible never to *renegotiate* in order to prevent a reversion to a price war? If the firms have the opportunity to communicate fully at time zero, why do they not again have such an opportunity in the event of a defection? Certainly a casual study of cartel behavior suggests that renegotiation occurs. While it is in the interest of the firms to commit themselves *not* to renegotiate, such behavior may not be credible. Indeed, to the extent that a defector knows he will be forgiven in the renegotiation, defection becomes attractive and collusion difficult. We see here a corollary of the topsy-turvy principle. Topsy-turvy antitrust policies that permit oligopolists to renegotiate and re-establish a collusive outcome in the event of a defection may in fact undermine tacit collusion and promote competition.

In a very interesting paper, Farrell and Maskin (1987) have developed a concept of renegotiation in repeated duopoly games. I think of their renegotiation concept as a type of "collective perfection requirement". Farrell and Maskin define a *set* of equilibria (and the corresponding payoffs) in a supergame as *renegotiation proof* if it satisfies two conditions: (a) each equilibrium in the set is supported as an equilibrium using only other members of the set itself as equilibria in subgames, and (b) no point in the set Pareto dominates any other point. The idea here is that the firms have identified the points in the renegotiation proof set as achievable, and will again find them achievable in any subgame, but they realize that it never is credible for the players as a group to pick a Pareto dominated continuation equilibrium. Implicit in this definition is that there will be no breakdown in the renegotiation process (such a breakdown could itself be a punishment if it is credible). In general, there will exist many renegotiation-proof equilibrium sets.

Farrell and Maskin demonstrate two conditions that characterize renegotiation-proof equilibria for discount rates close enough to unity: first, the player doing the punishing must prefer the punishment regime to the original equilibrium; and second, the player being punished must rather behave as specified in his punishment than defect yet again while his rival attempts to punish him. Typically, punishment consists of switching to an equilibrium (in the renegotiation-proof set) that is unfavorable to the defecting firm.

In their application of the renegotiation-proof concept to repeated oligopoly, Farrell and Maskin show how the requirement that equilibrium be renegotiation proof restricts the set of equilibrium payoffs that can be supported as supergame equilibria for discount factors close to unity. Whereas all points on the Pareto frontier are supergame equilibria for δ close to 1, those points giving the firms very different payoffs are not renegotiation proof: a firm with a very low payoff would simply defect, knowing that its rival would rather return to the original equilibrium than engage in punishment.

For the example of Cournot duopoly with $p(X) = 1 - X$ and zero costs, the Pareto frontier is the set of (flow) payoffs $(\pi, 1/4 - \pi)$, for $0 \leq \pi \leq 1/4$. In contrast, only points at which each firm earns at least $1/36$ each period are renegotiation proof. The set of renegotiation-proof equilibrium payoffs can be narrowed somewhat further by requiring that the renegotiation-proof set contain no points Pareto dominated by any element of *any* renegotiation-proof equilibrium set. Farrell and Maskin call such sets *strongly renegotiation proof*. In the example above, a payoff on the Pareto frontier can be sustained in a strongly renegotiation-proof way if and only if each player earns at least $1/16$ on a flow basis.

In view of the obvious incentives of oligopolists to avoid having to engage in mutually destructive punishments, I consider the concept of renegotiation-proof equilibria a promising way of refining supergame theory.[89] Unfortunately, although the renegotiation-proof criterion is appealing and narrows down the set of equilibria, a great many equilibria survive this refinement.

4. Two-stage competition

Repeated games are limited in that history has no tangible effect on prospective competition. The subgame beginning at any date is identical to the original game.

[89]See also the independent development by Bernheim and Ray (1987). Pearce (1987) has developed a rather different concept of renegotiation proofness and applied it to oligopoly theory, and several authors are currently working on notions of renegotiation, e.g. in the context of contracting.

The remainder of this chapter moves away from this very strong assumption to look at the wide range of competitive behavior involving investments that materially alter the subsequent competitive environment. Extending our study to such models allows us tó explore such concepts as pre-emption and strategic commitment.

The natural place to begin studying strategic behavior is in the context of two-period models of oligopoly. Two-period models have the advantage of being quite tractable while highlighting the importance of timing. They suffer, however, from their artificial timing structure: the second and final period is essentially one of static oligopoly, albeit one influenced by the first-period actions. For truly dynamic models of repeated interaction, see Section 5 below.

It is important to realize the essential role of sunk costs in a dynamic, strategic environment. All of the analysis to follow considers equilibria in games where the firms can make *commitments* at early dates that influence the competition to follow. But no action is a commitment if it is swiftly and costlessly reversible. It is the sunkness – at least the partial sunkness – of various investments that qualifies them as strategic decisions. Sunkness is implicit in all of the models I refer to in this section. For example, when I discuss the strategic aspects of investments in physical plant and equipment, it must be understood that such investments are strategic only to the extent that they would be costly to reverse, i.e. only to the extent that they are sunk investments. This proviso applies to virtually all of the recent literature on dynamic oligopoly theory, and should serve to emphasize one of the themes of the contestability literature, namely the importance of sunk costs.

An enormous number of two-period models of strategic interactions have been studied in the past several years. Each of these models employs the equilibrium concept of subgame perfection, requiring that the firms correctly anticipate the outcome of the second-period competition as a function of any first-period choices that they make. All of these two-period models conform to the following general structure: in the initial period, one or more of the firms has the opportunity to take some action that will have real economic consequences for the state of competition in the second period. Let us call these first-period actions the "strategic" ones. In the second period, some simple Nash equilibrium emerges, given the conditions inherited from the initial period. Subgame perfection requires that these second-period actions form a (static) Nash equilibrium in the game at that time. The second-period strategies might be called "tactical" responses to the earlier strategic choices.

The analysis invariably focuses on identifying the "strategic effects" that influence first-period behavior, and on attempting to characterize the resulting strategic rivalry. Isolating the strategic effects requires one to define some nonstrategic baseline for first-period behavior. In what follows, I take as the baseline the open-loop equilibrium in which each firm chooses its first-period actions taking as given its rivals' actions in *both* periods.[90] Any differences between the open-loop equilibrium and the subgame perfect equilibrium arise because a firm accounts for the influence of its first-period strategy on its rivals' later actions. Many of the basic ideas of strategy and tactics found in this growing literature can be traced back to Schelling's (1960) classic book, *The Strategy of Conflict*.

4.1. *A simple model of two-stage competition*

I begin my treatment of two-period models by outlining a very simple generic model of two-period duopolistic competition that shows the basic mathematics of strategic competition. Below, I shall show how a very large number of ideas in oligopoly theory can be interpreted as strategic and tactical behavior in two-period models.

The basic situation in which we are interested is as follows: a first-period decision by one firm has an effect on the environment in which rivalry is played out in the future, and hence on the subsequent choices made by that firm's rival or rivals. To isolate this strategic principle, and to show how it alters the original firm's first-period behavior, my generic model permits only one firm to make a strategic decision in period one.[91]

Let firm 1 have the opportunity to make a strategic investment during the first period. Measure the extent of firm 1's investment, or commitment, by its first-period outlay, K (measured in period-two dollars). For simplicity, assume that K alters the costs (or demand) faced by firm 1 during the second period, but does not affect firm 2's costs (or demand). The obvious example is investment in some type of capital (physical or human) that affects variable costs. After this investment is made, the firms play some noncooperative duopoly game in the second period. Writing firm 1's second-period profits as $\pi_1(x_1, x_2, K)$, where x_i is firm i's second-period action, firm 1's total profits (again, measuring in period-two dollars) are $\pi_1(x_1, x_2, K) - K$. The sunkness of the investment K is implicit in this specification, which does not permit any disinvestment during the second period. Firm 2's profits are $\pi_2(x_1, x_2)$.

To find the subgame perfect equilibrium in this game, one must first determine the equilibrium in the second period for *any* possible K (i.e. for any possible

[90] Formally, the open-loop equilibrium is a Nash equilibrium in strategy profiles, where a firm's strategy profile indicates what action it will take at each date that it moves. In an open-loop equilibrium, a firm cannot affect its rivals' future actions by its own current actions, since it takes as given both its rivals' current actions and their future actions.

[91] The principles emerging from my simplified structure are doubly present when each of several firms can act strategically. A natural class of games in which only a single firm can strategically invest are entry-deterrence games in which only one of the oligopolists, the initial incumbent, is active in the initial period. See Chapter 8 by Richard Gilbert in this Handbook.

first-period history), and then "fold back" to determine firm 1's optimal choice of K. Formally, we can think of K as a shift parameter in firm 1's profit function, and hence in its reaction function. At a cost, firm 1 can shift its own reaction function through its choice of K. Observe that the strategic aspect of the investment K in this model is *not* that it alters firm 2's incentives or opportunities, but rather that investment by firm 1 alters that firm's *own* incentives at a later date.[92]

For any given K, the equations determining x_1 and x_2 are simply $\partial\pi_i/\partial x_i = 0$, $i = 1, 2$. These two equations define the Nash equilibrium choices in the continuation game as a function of K, $x_i^*(K)$. The equilibrium second-period profits are therefore given by $\pi_i^*(K) \equiv \pi_i(x_1^*(K), x_2^*(K), K)$.

What governs firm 1's choice of K? Acting strategically, firm 1 sets K to maximize $\pi_1^*(K) - K$. Differentiating this with respect to K using the definition of $\pi_1^*(K)$ gives:

$$\frac{\partial\pi_1(x_1^*, x_2^*, K)}{\partial x_1}\frac{\mathrm{d}x_1^*}{\mathrm{d}K} + \frac{\partial\pi_1(x_1^*, x_2^*, K)}{\partial x_2}\frac{\mathrm{d}x_2^*}{\mathrm{d}K} + \frac{\partial\pi_1(x_1^*, x_2^*, K)}{\partial K} - 1 = 0.$$

The first term here is zero by the definition of x_1^*, so we have:[93]

$$\frac{\partial\pi_1(x_1^*, x_2^*, K)}{\partial x_2}\frac{\mathrm{d}x_2^*}{\mathrm{d}K} + \frac{\partial\pi_1(x_1^*, x_2^*, K)}{\partial K} = 1. \tag{30}$$

The two terms on the left-hand side of equation (30) measure firm 1's marginal benefits of increased investment. The second of these terms is the direct effect on firm 1's profits; more capital expenditures in the initial period lead to smaller (variable) costs during the second period. This effect of course has nothing to do with the presence of a competitor or the opportunity to act strategically. It is the first term in (30) that captures the strategic incentive to invest. In the case of Cournot competition, $\partial\pi_1/\partial x_2$ tells us how firm 1's profits vary with firm 2's output, and $\mathrm{d}x_2^*/\mathrm{d}K$ measures the effect of firm 1's investment on firm 2's output.

[92] Of course, if K appeared directly in π_2, then firm 1 *could* alter firm 2's incentives, but this channel is not necessary for firm 1 to act strategically.

[93] The fact that the first term drops out is an example of a general application of the envelope theorem to multistage games. Given that a firm is optimizing in the future, its actions today, while they may affect its own future choices, cannot (to the first order) thereby raise the firm's objective function. To alter the firm's subsequent payoff, today's actions must either have direct effects on its future payoffs or alter the actions taken by others in the future.

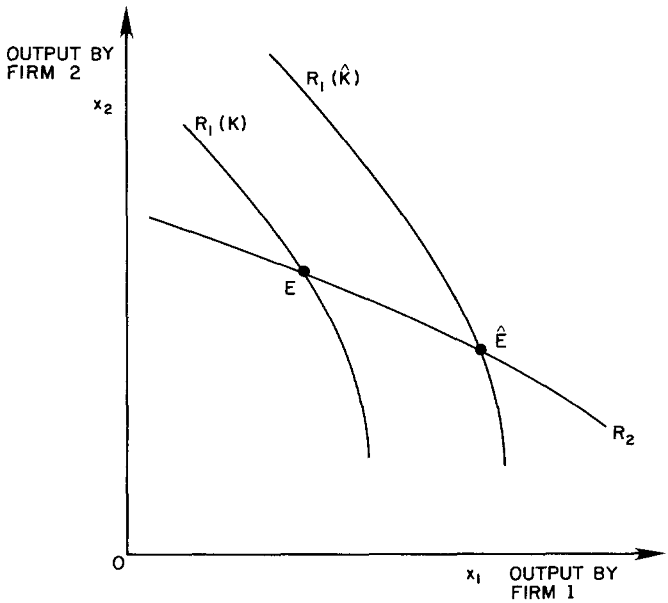

Figure 6.4. Strategic investment with quantity competition ($\hat{K} > K$).

How does K alter x_2? In the current simple model, where K does not directly enter into firm 2's profit function, K can only alter x_2 through its effect on x_1.[94] Again, the point is that firm 1 alters its *own* incentives in the second period, and thereby manipulates firm 2's behavior at that time. This "self-manipulation" is displayed in Figure 6.4 for the standard case of Cournot duopoly with homogeneous products. In that case, firm 1's reaction curve is shifted outwards by additional investment; this in turn causes firm 2 to contract, to firm 1's benefit. The conclusion is that, in a homogeneous-product Cournot oligopoly, a firm has a strategic incentive to invest in capital as a way of increasing its market share and profits.

More generally, but still within the context of our formal model, by working through the comparative statics effect of a change in K on x_1 and x_2, we have:

$$\frac{\mathrm{d}x_2^*}{\mathrm{d}K} = \frac{\partial^2\pi_1}{\partial x_1\,\partial K}\,\frac{\partial^2\pi_2}{\partial x_1\,\partial x_2}\Big/|M|, \tag{31}$$

[94]If K shifts the demand for firm 2's products or firm 2's costs, then it would *directly* influence x_2, and the formulas to follow would have an additional term.

where $|M|$ is the determinant of the second-derivative matrix:

$$M \equiv \begin{pmatrix} \dfrac{\partial^2 \pi_1}{\partial x_1^2} & \dfrac{\partial^2 \pi_1}{\partial x_1 \, \partial x_2} \\ \dfrac{\partial^2 \pi_2}{\partial x_2 \, \partial x_1} & \dfrac{\partial^2 \pi_2}{\partial x_2^2} \end{pmatrix}$$

associated with the system of equations $\partial \pi_i / \partial x_i = 0$, $i = 1, 2$. $|M|$ is positive if the system is stable [i.e. if inequality (5) is satisfied].

The preceding analysis shows that the sign of the strategic effect, the first term in (30), is the same as the sign of

$$\frac{\partial^2 \pi_1}{\partial x_1 \, \partial K} \frac{\partial^2 \pi_2}{\partial x_1 \, \partial x_2} \frac{\partial \pi_1}{\partial x_2}.$$

This expression is the product of (i) the effect of investment on 1's incentive to produce output, (ii) the effect of 1's output on 2's output, and (iii) the effect of 2's output on 1's profits. In the case of Cournot duopoly and physical capital investment discussed above and displayed in Figure 6.4, (i) is positive, while (ii) and (iii) are negative.[95] On net, firm 1 has a strategic incentive to invest because this causes it to produce more, and hence firm 2 to produce less, which is to firm 1's advantage.

It is a simple matter to apply the above model to differentiated product pricing competition instead of Cournot duopoly. Simply re-interpret the second-period strategies x_1 and x_2 as prices instead of outputs (with apologies for the notation). Now, added investment reduces 1's costs, and leads firm 1 to set a lower price, $\partial^2 \pi_1 / \partial x_1 \, \partial K < 0$, which induces firm 2 to lower its price, $\partial^2 \pi_2 / \partial x_1 \, \partial x_2 > 0$, harming firm 1, since $\partial \pi_1 / \partial x_2 > 0$. Consequently, firm 1 has a strategic incentive to *under*invest in capital, as a way of keeping prices high. This case of upward-sloping reaction curves is shown in Figure 6.5.

4.2. *Welfare effects of strategic behavior*

It is difficult to make general statements regarding the welfare consequences of the strategic behavior identified above. As usual in Industrial Organization, there are two wedges to consider: (1) the effect of the practice on consumers, and (2) the effect on rival firms. Since strategic investment by firm 1 is undertaken up to

[95]Strictly speaking, (ii) is negative only if b_2 from equation (4) is negative, as I assume.

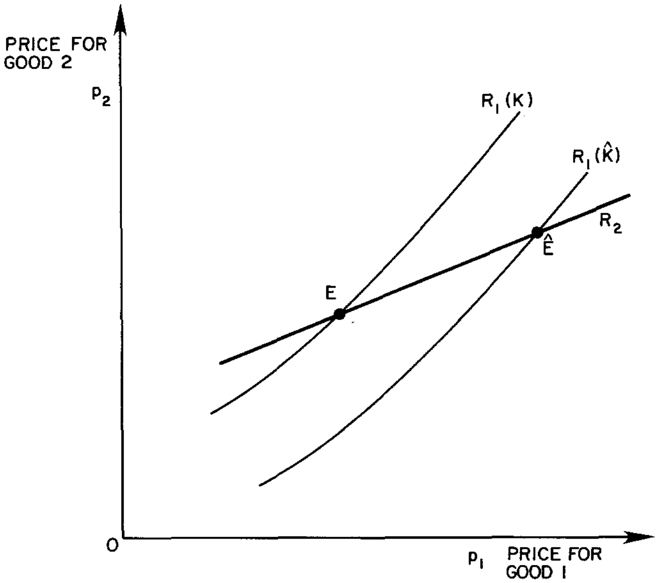

Figure 6.5. Strategic investment with pricing competition ($\hat{K} < K$).

the point where its marginal effect on firm 1's profits is zero, its marginal effect on welfare is the sum of its effects on consumers and on rivals.

In the case emphasized above – investment in physical capital followed by quantity competition – investments by firm 1 harm firm 2 (since firm 1 takes a more aggressive stance), but they typically *benefit* consumers since total output rises (here we are using comparative statics results from Subsection 2.1). The net effect cannot in general be signed: we cannot be sure whether the strategic incentive to invest actually raises or lowers welfare.[96]

The situation is quite similar in the case of strategic investment and pricing competition. Again the two wedges point in different directions: strategic under-investment by firm 1 leads to less fierce pricing competition, to the benefit of firm 2. But now consumers are made worse off on account of higher prices. As much as we have learned about strategic behavior, we unfortunately have little *general* guidance regarding the effect of this behavior on overall industry performance. In particular cases (see below) welfare results can be obtained, but the wide variety of strategic behavior precludes any general welfare theorems.

[96]Brander and Spencer (1983) is an example of this ambiguity in the context of R&D investment. Similar ambiguities arise when investments are used to deter entry.

4.3. Taxonomy of strategic behavior

There is a general point underlying the distinction between Figures 6.4 and 6.5. In the case of quantity competition, investment by firm 1 causes that firm to behave more aggressively (by selling more in response to any x_2), and this in turn induces firm 2 to behave *less* aggressively (actually sell less in equilibrium). Strategic overinvestment is the result. In the case of pricing competition, investment by firm 1 again leads it to be more aggressive, but in response to firm 1's lower price, firm two behaves *more* aggressively, not less, since it lowers its own price. This leads to strategic underinvestment.

This distinction between the downward-sloping reaction curves of Cournot rivalry and the upward-sloping reaction curves of Bertrand competition has been understood for some time. But recently, Fudenberg and Tirole (1984) and Bulow, Geanakoplos and Klemperer (1985) have supplied us with a taxonomy of such possibilities and some language to describe them. To establish a consistent notation, define the x_i's so that $\partial \pi_i / \partial x_j < 0$; a higher x_i represents more aggressive behavior by firm i.

Bulow, Geanakoplos and Klemperer have coined the terms "strategic substitutes" and "strategic complements" to capture the two cases just discussed (although this distinction was generally appreciated well before their paper was written). "Strategic substitutes" is defined by the inequality $\partial^2 \pi_2 / \partial x_1 \, \partial x_2 < 0$, while "strategic complements" is the reverse inequality. In their terminology, I have just shown that quantity competition involves strategic substitutes, whereas price competition involves strategic complements [assuming that the appropriate conditions, equations (4) and (12), are satisfied]. With strategic substitutes, more aggressive behavior by firm 1 (an increase in x_1) induces less aggressive behavior by firm 2 (a reduction in x_2). As Bulow, Geanakoplos and Klemperer point out, the normal notion of substitutes captures the effect of x_1 on firm 2's profitability, i.e. $\partial \pi_2 / \partial x_1 < 0$, whereas strategic substitutes captures the effect of x_1 on the *marginal* profitability of firm 2, $\partial^2 \pi_2 / \partial x_1 \, \partial x_2 < 0$. They focus on the $\mathrm{d} x_2^* / \mathrm{d} K$ term from equation (31), calling this the core of their paper (see p. 494).

Fudenberg and Tirole (1984) provide a somewhat less formal but more general taxonomy of strategic behavior in two-period games. Certainly, their terminology is more graphic. When $\partial^2 \pi_1 / \partial x_1 \, \partial K > 0$, they say that "investment makes the incumbent tough". With downward-sloping reaction curves, $\partial^2 \pi_2 / \partial x_1 \, \partial x_2 < 0$, and we have our familiar overinvestment result, which Fudenberg and Tirole call the "top dog" strategy. With upward-sloping reaction curves, they call the underinvestment the "puppy dog" strategy: firm 1 underinvests in order to be a less threatening rival. The other possibility is that investment makes firm 1 "soft". This is defined by the condition $\partial^2 \pi_1 / \partial x_1 \, \partial K < 0$. As we know from equation (31), this reverses the strategic effect. With downward-sloping reactions, we now get underinvestment by the incumbent in order to be more aggressive, which Fudenberg and Tirole call the "lean and hungry" strategy. If the reaction curves slope upwards, we have the "fat cat" strategy of overinvesting to be less aggressive, and hence inducing one's rival also to be less aggressive. Fudenberg and Tirole's four diet/animal strategies correspond to the four possible sign combinations of the two terms in the numerator of equation (31).[97]

4.4. Examples of two-period models

There are a remarkable number of applications of the basic strategic effects identified above. Although many of these examples involve strategic behavior by more than one firm, the general principles we have just seen continue to describe the essence of the strategic interactions. In all cases, it is essential not only that a firm make a strategic commitment, but also that this commitment be communicated to the rival. Absent credible communication, no strategic advantage can be obtained.[98]

(1) *Stackelberg leadership*. Stackelberg leadership is the granddaddy of two-stage models. It demonstrates the value of being the leader, now called a *first-mover advantage*. First, one firm selects its output; then the other firm follows with its own output choice. Formally, the leader's strategy is an output and the follower's strategy is a function specifying its output for each possible output by the leader. Stackelberg (1934) identified the leader's strategic advantage, anticipating as it does its rival's response. In terms of Figure 6.1, the leader (say, firm 1) selects its most-preferred point on the follower's (firm 2's) reaction schedule. Clearly, the leader must be at least as well off as he would be as a Cournot duopolist, since he has the option of setting his Cournot output (knowing that this will induce Cournot behavior by the follower). In general, the leader can earn higher profits than his Cournot profits.

The presence of a first-mover advantage carries over to Stackelberg pricing games with differentiated products. If two firms set their prices sequentially, again a firm can improve on its Bertrand payoff by strategically selecting its

[97]Fudenberg and Tirole's taxonomy also includes the distinction between entry-deterring and entry-accommodating behavior. Let firm 1 be the incumbent in the first period and firm 2 be the potential entrant. Now the *total* effect of K on π_2, $\mathrm{d}\pi_2^*/\mathrm{d}K$, and not just the marginal effect of K on x_2^*, comes into play. The reason is that firm 2 has an entry decision as well as a choice of x_2, i.e. firm 2's profit function is not concave in x_2, since it must bear some fixed entry costs if it selects $x_2 > 0$. In Dixit's (1980) model, overinvestment both deters entry and gives a strategic advantage if entry is not deterred. But in other models firm 1 may overinvest to deter entry, even though it would underinvest if entry were a foregone conclusion. The general principle is to cause one's rival to behave less aggressively given that entry cannot be deterred, but to reduce one's rival's profits as a way of deterring entry. See Gilbert, Chapter 8 in this Handbook, for further discussion of this point.

[98]Again, I refer the reader to Schelling (1960) for a pioneering discussion of the importance of communicating threats and commitments to one's opponent.

price, realizing that choosing a higher price will induce its rival to do the same. An exercise for the reader is to identify the Stackelberg point on Figure 6.2.

In today's terminology, Stackelberg and Cournot equilibria are each (subgame perfect) Nash equilibria, but to different games.[99] The difference is solely one of the timing of moves, but the importance of timing is one of the main lessons to be derived from the whole class of two-period games. The most durable criticism of Stackelberg's work, namely the exogenous specification of who is the leader and who is the follower, foreshadows a general criticism of simple timing models: the results depend upon the exact specification of the extensive form game being played, so one must be careful to justify this exogenous specification.

(2) *Strategic investment.* One or more firms has the opportunity in the first period to make capital investments that will influence its variable costs in the second period. I referred to this example most often in the development of the generic two-period model above. Investment is strategic, and pricing or output decisions are tactical. These models were originally developed by Spence (1977) and Dixit (1980) to study entry deterrence. With quantity competition, there is a strategic incentive to overinvest; this tendency is reversed with pricing competition. Investing in inventories can be studied in a similar fashion, although this analysis is complicated by the presence of storage costs. See Rotemberg and Saloner (1985) for a discussion of tacit collusion in the presence of inventories, and Arvan (1985) and Saloner (1986) for some examples of the strategic use of inventories as investments.

Strategic investment in capacity need not lead to overinvestment. Our generic two-period model tells us that with pricing competition (and, say, differentiated products) there will be a strategic incentive to underinvest. Gelman and Salop (1983) identified a similar strategy in the context of optimal entry strategies: a firm may succeed in entry by committing itself (if possible) to remaining small and thereby evoking a less aggressive response on the part of its rival. In this case Fudenberg and Tirole's (1984) term "puppy-dog ploy" seems especially appropriate.

(3) *Learning-by-doing.* When firms learn from experience, their first-period production levels affect their second-period costs, and hence competition at that time. The strategic incentives here are much like those above for investment in physical capital, except that investment consists of a larger first-period production level. With quantity competition, strategic concerns drive firms to produce more at early dates than they otherwise would. Of course, when examining the time path of output, this effect must be balanced against the natural incentive to

[99]Without the refinement to subgame perfection, any output by the leader giving him non-negative profits would be an equilibrium (much as any individually rational outcome is an equilibrium in the simultaneous-move output game if the firms' strategies are supply functions). All but one of these equilibria are supported by incredible threats on the part of the follower, however.

produce more in the second period when marginal costs are lower. See Fudenberg and Tirole (1983a) for the clearest two-period exposition.

(4) *Cost-reducing R&D.* Consider a model in which oligopolists first engage in a phase of R&D and later play an output or pricing game. The R&D investments affect second-period costs, possibly with spillovers; see, for example, Spence (1984). With quantity competition, a firm realizes that lower costs confer on it a strategic advantage. See Brander and Spencer (1983) for a clear exposition of this in a two-period model. The strategic effect stimulates the firms' R&D investments. Again, the reverse is true with pricing competition.

(5) *Network competition.* In the first period, competing firms build up customer bases, which have value in the second period because buyers benefit from being part of the crowd ("demand-side scale economies"). Katz and Shapiro (1986a, 1986b) study the intense first-period competition to attract customers when products are incompatible. They also examine the strategic effects of designing compatible products in order to reduce this intense early competition.

(6) *Patent licensing.* Initially, firms sign contracts to be licensees of new products or processes. Given the provisions of these contracts, the licensees then compete as oligopolists. Each firm recognizes that giving its rival access to a superior technology will put it at a strategic disadvantage (or eliminate a strategic advantage) in the subsequent oligopolistic rivalry.[100] But licensing may still occur if a mutually beneficial licensing contract can be constructed. See Katz and Shapiro (1985).

(7) *Advertising.* In the first period, advertising budgets are chosen. These expenditures influence second-period rivalry since advertising is durable. As Schmalensee (1983) has shown, a firm may enjoy a strategic advantage by underinvesting in advertising. Such a strategy may make that firm more aggressive in seeking new customers during the second period, since it has fewer loyal customers. In Fudenberg and Tirole's language, this is the lean and hungry look.

(8) *Information exchange.* In an uncertain environment with private information about demand or cost conditions, rival firms first have the opportunity to exchange (reveal) their information about industry conditions. Then they compete in the second period on the basis of that information. Again the incentive to reveal information is different for quantity than for pricing competition. And it depends upon whether the information regards demand or cost conditions. See Vives (1984) and Shapiro (1986).

(9) *Mergers.* First, firms have the opportunity to consolidate their operations by merging. Then the set of remaining entities plays an oligopoly game. The first-period merger decisions are strategic. Salant, Switzer and Reynolds (1983)

[100] This theme has been explored very nicely by Salop and Scheffman (1983) under the rubric "raising rivals' costs", They discuss various tactics by which a firm might reduce the efficiency of its rivals' operations. One such tactic is to withhold a superior technology or key input from one's rival.

show for the case of constant marginal costs that mergers are typically unprofitable under Cournot competition. Davidson and Deneckere (1985a) show that this result is reversed for pricing competition. The reason is that a merger causes the participants collectively to behave less aggressively; this is a benefit with pricing competition, but not with quantity competition. Perry and Porter (1985) and Farrell and Shapiro (1988) examine mergers in the presence of increasing marginal costs, i.e. when firms own some industry-specific capital.

(10) *Product selection*. A natural way of studying product selection is to treat it as a first-period choice, followed by second-period pricing competition. Oligopolists have an incentive to choose product varieties that are not too close as substitutes, in order to diminish second-period pricing or output competition. For analyses along these lines see Prescott and Visscher (1977), Shaked and Sutton (1982), Brander and Eaton (1984), and Judd (1985).

(11) *Financial structure*. Another way in which a firm might commit itself to more aggressive behavior is through its choice of a debt–equity ratio. This possibility is explored in Brander and Lewis (1986). They consider a model in which firms first decide how much debt to issue (total capital requirements are fixed), and then compete as duopolists. The key point is that a more leveraged firm will act more aggressively in the output market, so long as profits are uncertain and bankruptcy is a possibility.

Debt financing leads to more aggressive behavior because of the following two effects: (i) with more debt, the owners of a firm have an incentive to increase the variability of its profits, and (ii) an increase in output raises the variability of a firm's profits in the presence of uncertain demand. The second point requires an assumption that the marginal profitability of output is higher when conditions are more favorable, i.e. $\partial^2\pi_i/\partial x_i\,\partial\theta > 0$, $i = 1, 2$, where overall profitability increases with the random variable θ. Point (i) follows from the observation that the returns to the owners are given by the maximum of π_i and 0; since this maximum is a convex function of π_i, a mean-preserving spread in π_i raises the returns to the owners. Put differently, the owners prefer large to small (positive) profits, but are indifferent between small and large losses. In particular, an increase in debt financing by firm i raises the critical value of θ_i, $\hat{\theta}_i$, at which the owner's of firm i break even (for a given choice of x_1 and x_2). To see point (ii), note that an increase in output (in the relevant range) raises profits in favorable states and lowers them in unfavorable states, so long as $\partial^2\pi_i/\partial x_i\,\partial\theta > 0$. These observations led Brander and Lewis to conclude that firms have a strategic incentive to use debt financing as a way of committing themselves to be more aggressive. Of course, this result is limited to the case of "strategic substitutes", and is reversed for second-period pricing competition.[101]

[101] Brander and Lewis (1985) explore another aspect of the strategic role of financing – the effect of bankruptcy costs – in their companion piece.

(12) *Labor contracts and managerial incentives*. Just as investment in physical capital can lower marginal costs and serve a strategic role, so can labor contracts or managerial incentive schemes.

In arranging its labor contracts, the key (with, say, quantity competition) is for an oligopolist to turn variable costs into fixed costs, thereby committing itself to a more aggressive stance. This could be done by raising base wages in order to reduce overtime rates (so that during a boom in demand an expansion in production calling for overtime work would be less expensive on the margin). Or a firm could agree with its union that the firm could hire new or temporary workers at wages lower than those paid to current union members. Similarly, a contract specifying generous severance pay would lower the opportunity cost of keeping workers on the job and induce more aggressive behavior during downturns in demand.

In designing managerial incentive schemes, the owner of an oligopolistic firm will strive to give incentives to the firm's manager to act aggressively (again, in the case of quantity competition). It should be apparent to the reader by now that allowing the owner of just one firm to commit itself to a managerial incentive scheme (e.g. paying the manager on the basis of market share, revenues, etc.) allows that firm unlimited commitment power. Consider the effect of such a commitment prior to a Cournot duopoly game. By appropriately rewarding the manager on the basis of sales, say, the owner with commitment power could achieve the Stackelberg leadership point. The more interesting case arises when all firms have equal opportunities to commit themselves to managerial contracts. As we would expect, in the case of quantity competition, each owner causes his manager to act more aggressively than otherwise, with an end result of more competition and lower profits. The reverse results hold for pricing competition. In either case, although it may appear to an outside observer that the managers are not maximizing profits (in the second stage), this is only because owners, in the interest of profit maximization, have strategically instructed them not to do so. See Fershtman and Judd (1984) and Sklivas (1985) for analyses along these lines. But as Katz (1987) shows, the strategic scope for incentive contracts is completely eroded if it is impossible to verify the existence of such contracts to one's rivals.[102]

(13) *Long-term contracts with customers*. There are many contracts that a firm might sign with customers in the first period in order to give it a strategic edge in

[102] In particular, effective commitment by an owner–manager pair requires that the owner and the manager be able to convince others that they have not written a *new* contract superceding the one to which they claim to have committed. The problem arises because, given what others believe firm i's contract to specify, it is in the joint interests of the owner and the manager at firm i to sign a new contract instructing the manager to maximize profits (given the actions of rival firms as induced by the contract they believed to apply). Katz makes a strong case that it is difficult to verify the primacy of any contract that the owner and manager reveal.

later competition. With declining marginal costs, for example, a firm that has already contracted for a significant amount of business will have lower marginal costs for the remaining business and act more aggressively. Or a firm could offer first-period customers a discount if they are left stranded with an unpopular product, i.e. one with a small final market share (this would give the firm an incentive to achieve a high market share in the second period).[103]

Another example that I have discussed in Subsection 2.4 above is that of Retroactive Most-Favored-Customer Clauses. As Salop (1986) and especially Cooper (1986) note, a firm can alter its second-period incentive to cut prices by signing most-favored-customer agreements with first-period customers. Given these contracts, the firm can commit itself to less aggressive behavior, a benefit (and hence a facilitating practice) in a subsequent Bertrand game.

Holt and Scheffman (1986) show how most-favored-customer clauses and meet-or-release provisions of contracts can support the Cournot equilibrium (as well as more competitive equilibria) in a two-period pricing model. Each firm, realizing that second-period discounts from the first-period list price will be matched, takes other firms' outputs as given, just as in the Cournot model. A price reduction can increase a firm's sales by attracting unattached buyers, but since it cannot cut into rival sales, it acts exactly like an increase in quantity in the single-period Cournot model. Holt and Scheffman also consider *selective* discounts, whereby a firm offers a lower price only to a particular group of its rivals' buyers. They find that when firms can offer selective discounts, it is generally not possible to support supra-competitive list prices.

(14) *Investing in disinformation.* So far I have emphasized first-period actions that influence the state of second-period rivalry through tangible variables such as capacity or product selection. Another way in which first-period decisions can influence later behavior is through the *information* that they generate. Needless to say, these effects can only arise when the firms are uncertain about some aspect of the market environment, and when one firm's actions affect the information received by its rival(s) about the market.

Riordan's (1985) model is a nice example of strategic manipulation of information in a duopoly.[104] The two firms play a twice-repeated Cournot game, but are uncertain of the demand conditions in either period and cannot observe their rival's output level. Critically, the demand shock is positively serially correlated across the two periods. Each firm uses the first-period price as information in its effort to estimate the demand shock at that time. A lower price in the first period

[103]Contracts in which the price ultimately paid is contingent on the firm's market share are natural in the presence of network externalities, because customers then care directly about the market share of the brand that they buy.

[104]Strategies of this type have been explored more fully in the context of entry deterrence. See Chapter 8 by Richard Gilbert in this Handbook, the papers on limit pricing by Milgrom and Roberts (1982a) and by Harrington (1986), and that on predation by Kreps and Wilson (1982b).

leads a firm to estimate that demand was less favorable. This in turn – due to the serial correlation of demand shocks – convinces the firm that second-period demand is weaker, and the firm reduces output accordingly. So, a lower first-period price induces each firm to select a lower second-period output. But now note that each firm has an incentive to manipulate the other's information: by producing more at the beginning, a firm can cause its rival to take a more pessimistic view of market conditions and produce less later on (hence the title of the paper). Again we get overinvestment in the first period: the firms produce more at that time than they would under static Cournot competition. This result would be reversed if demand shocks were negatively correlated across periods, due, for example, to the presence of inventory holdings by buyers. An opposite result also would be found if the firms were Bertrand rather than Cournot competitors. As with so many of the results in this section, the strategic incentives to manipulate information are sensitive to the precise nature of the tactical competition.

An analogous overproduction result also arises in a model with cost uncertainty. Suppose that each duopolist is uncertain of its rival's cost, and they play a twice-repeated Cournot game. Now each firm has an incentive to expand its production during the first period in an effort to make its rival believe that its costs are lower than they in fact are. Such a message would cause the rival to scale back its second-period output. Mailath (1985) analyzes the corresponding duopolistic two-period pricing game with uncertain costs. He finds that cost uncertainty causes the firms to behave less competitively, as each firm would like its rival to believe its costs are high. See Roberts (1987) for an excellent and much more complete discussion of models of this sort, and Milgrom and Roberts (1987) for an insider's critique of the contribution of the literature on asymmetric information games to industrial organization.

(15) *Customer switching costs*. If consumers must bear a cost when switching from one supplier to another, the first-period sales have a lasting effect on competition. As with network competition and learning-by-doing, each firm benefits in the second period if it acted aggressively during the initial period. In many cases, however, having a base of locked-in, first-period consumers may make a firm less aggressive during the second period, because the firm's incentives to lower its prices are reduced (assuming that it cannot price discriminate). See Klemperer (1985) and Farrell and Shapiro (1987).

(16) *Multimarket oligopoly*. Commitments made in one market can affect subsequent competition in a second market, if there are linkages between markets on either the cost side or the demand side. Naturally, these ideas have application to international oligopolistic competition. Increased involvement by a given firm in market A, due to some commitment by the firm to that market, will cause that firm to behave more aggressively in market B if either (i) there are economies of scale or scope across the markets, so that the firm's marginal costs in market B are lowered by its increased presence in market A, or (ii) there are positive

demand interdependencies across markets, so that selling in market A enhances the demand for the firm's products in B. In general, if the two markets are linked in either way, the firm must account for the strategic effects in market B of its (prior) presence or scale of operations in market A. See Bulow, Geanakoplos and Klemperer (1985) for a further development of this theme.

(17) *International oligopoly*. In the case of international oligopoly, the first-period strategic behavior may not be undertaken by the oligopolists themselves, but rather by their home governments acting in their own interests. Take the case of a home firm and a foreign firm competing for business in a third country. One might think that there was no constructive role for the home government, since its interests coincide with those of the home firm (to maximize that firm's profits). This would indeed be the case in the absence of strategic effects (e.g. for a domestic firm selling as a monopolist in a foreign market). But the home government may be able to induce the home firm into more aggressive behavior, even when the firm could not so commit itself. For example, an export subsidy to the firm would have this effect. With quantity competition, therefore, the increase in profits can exceed the subsidy payments, generating a net domestic gain at the expense of the foreign competitor. See Brander and Spencer (1985) for a clear exposition of this case, and Krugman (1984) and Dixit and Kyle (1985) for related analyses.

While some policymakers cherish this result as a justification for their protectionism, it should be applied with extreme caution for at least three reasons. First, its application can easily lead to an "export subsidy war". If each country can subsidize its home firm, the ultimate beneficiaries will be consumers in the third country. Just as with strategic investment, we have a prisoner's dilemma structure: if each government acts strategically, both exporting countries are worse off. Second, the policy prescription is sensitive to the assumption of quantity competition, i.e. to the case of "strategic substitutes". If the firms play a pricing game, export taxes rather than subsidies are optimal (although the "export tax war" benefits both countries, in contrast to the subsidy war under quantity competition). See Eaton and Grossman (1986) for an explanation of when taxes are optimal and when subsidies are optimal. Third, Dixit and Grossman (1986) point out that a subsidy to one industry must ultimately come at the expense of some other industry, and there may be a decrease in its profits if it, too, is oligopolistic. In a symmetric model, subsidies are not optimal, even when firms play quantity games.

In this section I have explored a great many dimensions of strategic behavior. It should be clear that strategic considerations may encourage or discourage investments of various sorts, depending upon the specifics of the post-investment rivalry. But we do have a unified theory in the sense that the same strategic principles apply in so many economic environments, and we can generally relay on the comparative statics properties of our *static* oligopoly theory to provide information about strategic behavior. What we lack – and should not strive for, in my opinion – is a general theory of the effect of such strategic behavior on the firms' profits or on industry performance. In some cases, strategic investments heighten competition to the firms' disadvantage (a prisoner's dilemma structure at the investment stage) and consumers' advantage. In other cases, quite the opposite is true. Again, we must look in some detail at the characteristics of a particular industry if we are to determine into which group it falls.

5. Dynamic rivalry

I have yet to discuss truly dynamic models of oligopoly – models of many periods in which the economic environment changes with time. In other words, I have not combined the repeated rivalry aspects of supergames with the investment and commitment aspects of two-period models. That is the topic of the current section. Naturally, such games are the most complex of all those discussed in this chapter, and their tractability represents a genuine problem. But they hold out the most hope of advancing our understanding of oligopolistic rivalry, and currently represent the area of greatest research activity in oligopoly theory. For the reader planning to pursue independent research in this area, I recommend close study of Fudenberg and Tirole's (1986a) excellent monograph on dynamic oligopoly in addition to the material below.

5.1. What makes for dynamics?

If we are to study dynamic oligopoly games, we had best understand what the fundamental sources of these dynamics are. In other words, why does an oligopolistic market environment change over the course of time? It is useful to categorize the sources of dynamics for the purposes of developing different models and theories.

Some industries experience changing conditions quite independently of their own behavior. One obvious example is an emerging industry where, say, exogenous technological progress is rapidly lowering costs; such industries are the subject of the literature on adoption of new technology (see Chapter 14 by Reinganum in this Handbook). Another example would be a declining industry where, say, demand is falling over time [see, for example, Ghemawat and Nalebuff (1985)]. Or an industry may experience cyclical changes in demand or factor cost conditions, perhaps due to economy-wide business cycles. A thorough treatment of these *exogenous* sources of industry dynamics, while a rich topic in the study of oligopolistic behavior, is beyond the scope of this chapter.

Oligopolistic markets may also experience *endogenous* changes in the conditions of competition. These endogenous changes, being subject to strategic maneuvering by the oligopolists, are the subject of this section. I will distinguish two types of industry conditions that may evolve over time. The first are *tangible* industry conditions such as the firms' capital stocks, the firm's technological capabilities, or previous commitments made by buyers to particular sellers. See Section 4 for more examples of tangible capital variables that may change over time in response to firms' strategic choices. I shall emphasize these tangible industry conditions below. The second type of industry condition that may be strategically controlled is an intangible, i.e. the *beliefs* about market conditions that are held by the oligopolists or their customers. The point is that one firm may take actions designed to manipulate its rivals' information, and hence their future actions, in a way that is favorable to itself. The circumstances in which this signaling behavior can be expected to arise, and the mechanisms through which it operates, are quite different from strategic investments in tangible assets. For these reasons, the literature breaks quite naturally into these two categories.

5.2. *The state–space approach to dynamic oligopoly models*

In turning our attention to models of dynamic oligopoly, we need to refine our methods in order to focus our attention on the strategic aspects of commitment, just as we already have done in Section 4 for two-period models. The particular problem that we immediately face is the following: if we are to examine infinite horizon models, all of the intricacies that were present in the study of supergames remain with us in principle. For example, consider a dynamic game in which the oligopolists make investment decisions each period that affect their capital stocks, and then play a pricing or output game each period given the associated variable cost curves. In such a game, the whole supergame calculus of defection and retaliation remains present; indeed, it is made more complex by the fact that the gains from defection and the credibility of punishment are dependent on the firms' capital stocks.

In order to focus on the strategic aspects of competition, i.e. on the changing economic environment, it is extremely useful to fold the tactical decisions into the background and work with some *reduced-form profit functions* indicating the firms' flow profits as functions of the "state variables". The state variables measure the economic conditions at any point in time. Formally, a state variable is anything that affects the ensuing subgame. Typically, the state variables adjust slowly in response to the firms' current decisions. In the example of an investment and pricing game, the state variables are simply the firms' capital stocks, and we would assume that firm i's profits during period t are given by some function $\pi_i(\boldsymbol{K}_t)$, where $\boldsymbol{K}_t \equiv (K_{1t}, \ldots, K_{nt})$ and K_{it} is firm i's capital stock at date t.[105] Using the language of Section 4, the firms behave strategically by making medium- to long-run commitments that alter market conditions (the *state*) into the future; short run, tactical decisions are subsumed in the reduced-form profit functions.

Restricting attention to strategies that depend only on the state of the industry is exactly the *opposite* of the approach taken in the supergame literature. In supergames, there is *no* state variable, so the state–space restriction would require the firms' strategies to be constant over time. Simple repetitions of the static, stage-game equilibrium would be the only possibility. And in dynamic games, the most collusive equilibria (which were the focus of the analysis in supergames) generally rely on strategies that are *not* merely functions of the tangible state variables. These observations may make it appear that the state–space approach is very restrictive. But all it really does is rule out the type of bootstrapping that led to so many supergame equilibria. To see this, note that in a finite horizon game (with a unique equilibrium) the strategies necessarily are functions only of the state of the market.[106] All of the two-period models discussed above were state–space models in this sense. What the state–space approach allows us to do is focus on dynamic strategies involving commitment, i.e. those that alter future market conditions, without restricting our attention to finite-horizon games.[107] Another advantage of the state–space approach, in contrast to supergame theory, is that the equilibrium structure generally does *not* change suddenly as the number of periods becomes infinite.

In the case of tangible capital variables, the capital stocks themselves serve as the state variables. Anything that involves a commitment – either by one of the oligopolists or by one of their customers – can serve as a state variable, since commitments affect the prospective competition, i.e. the continuation game. I have already given many examples of state variables that apply in different markets. For these tangible variables, the equations of motion of the state variables are determined by the technologies of production and consumption, e.g. by the sunkness of investment made by each oligopolist. In the case of intangible variables, the state measures the firms' or consumers' beliefs about relevant but uncertain market conditions (such as cost or demand parameters). These models are quite different in structure, since the equations of motion for the state

[105] We could derive the $\pi_i(\boldsymbol{K})$ functions from the underlying cost and demand conditions by assuming, say, Cournot behavior at any date.

[106] This follows from our use of subgame perfection as an equilibrium concept. Previous behavior (cooperation vs. defection) is critical in determining firms' current behavior in the supergame literature, but past actions cannot affect current behavior in state–space games unless they alter current conditions.

[107] Finite horizon games often are awkward to analyze because the number of remaining periods is ever-changing. A finite horizon introduces an *exogenous* element of nonstationarity into the model; calender time is one of the state variables. Many of the lessons of finite-horizon state–space games can be gleaned from two-period models.

variables are determined by Bayesian updating, given some priors as initial conditions.

In adopting the state–space approach to dynamic oligopoly, we are implicitly putting a great deal of weight on the *payoff-relevant* features of the economic environment, i.e. those features that enter directly into the firms' prospective payoffs. In practice, there are a great many payoff-relevant capital variables in a given market; the art here is to identify the qualitatively important ones for the purpose of industry analysis. The multitude of possible strategic variables highlights both the importance of strategic commitments and the criticality of identifying the most significant strategic variables in any particular industry. As with two-period models, the literature is most naturally partitioned according to the particular state variables being studied. I follow this organizing principle for the remainder of this section.

5.3. *Pricing and quantity games*

It is natural to begin with our old friends, prices and quantities, as candidate strategic variables. Although prices are most often tactical choices folded into the reduced-form profit functions (as in the investment and pricing example outlined above), there are a number of settings in which they themselves can serve as the state variables. Previous pricing and production decisions *directly* influence current and future rivalry when these variables are sluggish or costly to adjust. In such circumstances, prices or quantities exhibit inertia, and hence have a direct effect on the continuation game.[108]

Cyert and DeGroot (1970) were the first to explore the commitment role of production decisions in an explicitly dynamic game. They examined an *alternating-move* duopolistic quantity game. With alternating moves, in contrast to the repeated but simultaneous move structure of supergames, there is a genuine state of the market when firm 2, say, is on the move. The state consists of firm 1's previous production choice, which by assumption must remain in force until firm 1 again has the opportunity to move. Cyert and DeGroot realized that by examining a finite-horizon, alternating-move quantity game, they could introduce genuine dynamics into the standard Cournot model without facing the problems associated with finitely or infinitely repeated games. Since the firms' commitments expire at different times, they actually are reacting to each other over time.

The alternating-move quantity game is a finite, perfect information game, so it (generically) has a unique perfect equilibrium. But the dynamics in this game's

[108]Of course, there are strategic aspects to pricing decisions whenever today's prices or outputs influence future competition through *other* state variables. For example, in the case of learning-by-doing, the state variable is cumulative sales, which is influenced by current prices. I return to these other state variables below.

perfect equilibrium are *not* trivial as they are in a finitely repeated Cournot game. Rather, we find a series of strategic commitments by the firms: each firm realizes when it comes to select a quantity that its choice will affect its rival's subsequent decision. Since these reaction functions are negatively sloped in the quantity game, each firm has an incentive to produce more than the Cournot duopoly output, much as we saw overinvestment in, say, Dixit (1980). Restricting their attention to quadratic payoff functions (such as would arise with linear demand and constant marginal costs), Cyert and DeGroot show that each firm's reaction function is linear. They use numerical methods to examine the limiting properties of the reaction functions as the finite horizon becomes long. In their example, the limiting behavior is indeed more competitive than Cournot behavior.

More recently, Maskin and Tirole (1987) have examined this same alternating-move game for more general profit functions and using an infinite horizon. They use the term *Markov perfect equilibrium* for the state–space approach, since the state space consists exactly of the rival's last quantity choice.[109] After displaying the general differential equations that the equilibrium reaction functions, $R_1(x_2)$ and $R_2(x_1)$, must satisfy, they too restrict attention to quadratic payoff functions. Looking for symmetric equilibria in linear reaction functions, they are able to prove that there exists a unique linear Markov perfect equilibrium, that this equilibrium is dynamically stable (i.e. from any starting point the output levels converge to the steady state), and that this equilibrium is the limit of Cyert and DeGroot's finite-horizon equilibrium.[110] They also establish that each firm's steady-state output is strictly greater than the Cournot output, unless the single-period discount factor, δ, equals zero (in which case the naive Cournot adjustment process is optimal, and steady-state behavior is simply Cournot). In fact, the steady-state output increases with δ. It is of some interest to note that the limit of the alternating-move game as the time between moves becomes small ($\delta \to 1$) does *not* approach the repeated Cournot outcome of the (finite-horizon) simultaneous-move game.

As we might expect, the more important is the future (large δ), the greater incentive each firm has to pre-empt its rival by expanding output; the equilibrium consequence is lower per-period prices and profits than under Cournot behavior. In this sense Maskin and Tirole's analysis provides support in an explicitly dynamic model for strategic principles we discovered in two-period models. But beware: what was technically a very simple problem in the two-period model – the

[109] See the first paper in their series, Maskin and Tirole (1988a), for further discussion of the Markov perfect equilibrium solution concept. They point out, for example, that a Markov perfect equilibrium is also a perfect equilibrium: given that other firms react only to the state variables, a single firm can also ignore the non payoff-relevant aspects of history.

[110] Here we have an example of why infinite horizon models can be easier to work with than their finite horizon counterparts: with the finite horizon, the reaction functions must be of the form $R_{it}(x_j)$; firm i's reaction depends not only on j's last move, x_j, but also on the number of periods remaining, hence t. With an infinite horizon, by contrast, we can write simply $R_i(x_j)$.

overinvestment result in the case of "strategic substitutes" – has become quite difficult in the infinite horizon model, and indeed has only been established in the very special case of linear demand and constant marginal costs. Note also that the assumption of alternating moves is somewhat artificial; this becomes all the more apparent if one considers more than two firms: oligopoly is not a board game.[111]

Just as we made the distinction between repeated quantity and repeated pricing games in Section 3, we can study alternating-move games in which the duopolists set prices rather than quantities. Maskin and Tirole (1986b) provides an excellent analysis of Markov perfect equilibria in the alternating-move pricing game with homogeneous products and constant marginal costs. Now each firm's strategy specifies a price that it will set in response to its rival's price: $R_1(p_2)$ and $R_2(p_1)$.

Unlike their unique symmetric equilibrium in the alternating-move quantity game, Maskin and Tirole identify multiple equilibria in the pricing game. They establish by means of examples that both Edgeworth cycles (see Subsection 2.2) or kinked demand curves (see Subsection 2.4) can be supported as Markov perfect equilibria. In the Edgeworth cycle equilibrium, each firm undercuts the other until prices are driven to marginal cost, at which point one of the firms (with positive probability) raises its price (to the monopoly price).[112] In the kinked demand curve equilibrium, each firm would more than match a price reduction (making such behavior unprofitable), but would not respond to a price increase. The equilibrium strategies here too involve mixing, but now mixing does not arise in equilibrium. Maskin and Tirole go on to characterize the class of kinked demand curve equilibria and to prove that for δ close to unity, an Edgeworth cycle equilibrium must exist. For δ close to unity, they also show that the monopoly pricing kinked demand curve is the only renegotiation-proof Markov perfect equilibrium (see Section 3 above for a discussion of the renegotiation-proof criterion).

We might expect on the basis of our two-period models that strategic pricing behavior would be quite different from strategic output choices, since each firm now recognizes that a higher price on its part will lead its rival to set a higher price in the subsequent period ("strategic complements"). Dynamic pricing reactions should lead to more collusive behavior than the static pricing game, although the reverse was true in the quantity game. Maskin and Tirole show that the intuition from two-period models does indeed extend – with some complications – to the infinite horizon model. In the quantity game, the cross partial derivatives of the profit function, $\partial^2\pi_i/\partial x_i\,\partial x_j$, are negative, leading to

[111]But Maskin and Tirole make significant progress in endogenizing the timing of the moves in their series of papers (1987, 1988a, 1988b).

[112]As Maskin and Tirole point out, these cycles do not rely on capacity constraints, as did Edgeworth's (1925) original example.

dynamic reaction functions that are negatively sloped and behavior that is more competitive than Cournot. In the pricing game, the comparable cross partials, $\partial^2\pi_i/\partial p_i\,\partial p_j$, vary in sign depending upon p_i and p_j, so the dynamic reaction functions are nonmonotonic; this is the source of the multiple equilibria. But in general, the Markov perfect equilibrium profits of the firms are positive, as opposed to the zero profits in the static Bertrand model with constant and equal marginal costs. And an increase in δ, i.e. more rapid responses, leads to more competitive behavior in the quantity game, but may well allow the monopoly price to be supported in a kinked demand curve equilibrium in the pricing game. This comparison of price and quantity games is encouraging insofar as it suggests that the findings in the two-period model are robust. By contrast, remember that there was virtually no distinction between pricing and quantity competition in supergames with δ near unity.

Another reason why prices or quantities might serve as commitments is that they may be costly to change. For the case of quantities, I defer this to the following subsection on capacities and investment games. Analyses of dynamic pricing games when there is a "menu cost" of changing prices have been developed by Marshak and Selten (1978) and Anderson (1985). both of these papers study games in which each firm can react very quickly to its rivals' price changes, but must incur a (small) cost every time it changes its own price. Anderson shows how collusion can be supported using simple strategies that match price cuts but not price increases (as in the kinked demand curve theory of oligopoly), even when the costs of changing prices are very small. More recently, Gertner (1985) has studied the Maskin and Tirole alternating-move model under the assumption that each firm bears a cost whenever it changes its price. This modification increases the number of state variables from one to two: even when firm 1 is on the move, its previous price (as well as firm 2's previous price) matters, because firm 1 would bear a cost if it were to set any other price. Gertner assumes that the firms can react quickly enough to each other's price changes so that no firm could benefit from a price change solely on the basis of its profits prior to its rival's reaction. He then establishes that the monopoly price can be supported as an equilibrium, although this is not the unique equilibrium. He also argues that the renegotiation-proof criterion can be used to select the monopoly price equilibrium.

5.4. *Investment games*

Much of my discussion of prices vs. quantities, the importance of dynamics, etc. suggests that we should study dynamic oligopoly when the firms make strategic investment decisions (with the state variables being physical capital stocks), and tactical pricing decisions. We did analyze two-period models of this sort in Section 4 above; it would be very useful to extend these models to many periods.

Much of the literature on dynamic investment games aims to understand the role of physical capital investments in deterring entry. The main references here are Spence (1979), Eaton and Lipsey (1980, 1981), and Fudenberg and Tirole (1983b). See Chapter 8 by Gilbert in this Handbook for further discussion of these games.

Rather few papers explore dynamic oligopolistic interactions among established firms who can make physical capital investments. One such model is Gilbert and Harris (1984), which investigates the time pattern of investments in a growing market. Gilbert and Harris assume that investments are "lumpy", i.e. come in discrete plant units. Given the state variable, k, the number of plants that have been built, each firm earns a flow of profits per plant of $p(k)$.[113] They look for the perfect equilibrium of investment dates. Since they assume that a plant is infinitely durable and allows production at constant marginal cost up to capacity, they effectively are looking at dynamic quantity choices, where quantities can only rise over time. Of course, infinitely durable capital permits firms to make significant, lasting commitments. Gilbert and Harris show that in a perfect equilibrium the race to install the next plant implies that it will be built as soon as market conditions are favorable enough so that it can earn non-negative profits. Of course, these profits must be computed over the entire, infinite lifetime of the plant, accounting for all future construction. This analysis shows how capacity competition can dissipate profits; but the paper does not give us much guidance on how investments affect tactical oligopoly pricing behavior, since the firms are assumed always to produce at capacity.

Benoit and Krishna (1987) and Davidson and Deneckere (1985b) examine the effect of investments on pricing strategies in a model where capacity choices are strategic and pricing competition is of the Bertrand–Edgeworth variety. Basically, they extend the two-period model of Kreps and Scheinkman (1983) to many periods.[114] In a duopoly model where the firms make once-and-for-all capacity choices and then compete via prices, we know from the Kreps and Scheinkman paper that it is an equilibrium for the firms to choose their Cournot quantities as capacities and then set the Cournot prices in each subsequent period; this is a trivial extension of the two-period model. But the point of extending the pricing game to many periods is to identify *other* symmetric equilibria.

[113] This reduced-form profit function would come about if, say, each plant had a unit capacity and involved no variable costs, and if the tactical oligopoly behavior involves production at capacity for all plants.

[114] Davidson and Deneckere use the same rationing rule (high value consumers buy at the lower price), and both papers use the same production technology (a firm with capacity x can produce up to x at no cost, but cannot produce more than x) as do Kreps and Scheinkman. Benoit and Krishna assume for some of their results that the demand function is concave, $p''(X) < 0$, whereas Davidson and Deneckere take demand to be linear.

Given initial investment levels $\{x_1, x_2\}$, we have a price-setting supergame with capacity constraints. So this analysis also extends Brock and Scheinkman's (1985) paper, which examined the pricing supergame under the assumption of exogenously given and equal capacities (see Section 3 above). Benoit and Krishna are able to establish the existence of equilibria involving higher prices than under Cournot behavior. By replacing the one-shot pricing game in the second period of Kreps and Scheinkman's two-period model with a pricing supergame, more collusive pricing equilibria are possible. This in turn has implications for the capacity decisions in the initial period. In particular, all equilibria other than the "Cournot" one involve the firms choosing capacities in excess of their subsequent production levels.[115] Both Benoit and Krishna, and Davidson and Deneckere, stress the result that the firms build excess capacity in order to better discipline their subsequent pricing behavior.[116] Firm 1 must build some excess capacity if it is to be in a position to punish firm 2 for expanding *its* capacity. In this sense, it is costly for the firms to sustain outcomes more collusive than Cournot.

It is also possible to wed capacity choices with the alternating move pricing games discussed in the previous subsection. Maskin and Tirole (1986b) find examples with small capacity costs and a discount factor near unity in which the firms build excess capacity that is never used. Gertner (1985) also identifies a strategic role for excess capacity in his model of price inertia.

5.5. *Intangible state variables*

In markets where the oligopolists face considerable uncertainty about underlying conditions, there is a potential role for strategic information manipulation or signaling. The state variables are the firms' or consumers' posteriors regarding such relevant variables as cost or demand parameters. Firms may invest, in the sense of sacrificing current profits, in order to manipulate their rivals' beliefs in a favorable way. In such games, it is important to specify the process by which beliefs are updated, both in and out of equilibrium (just as *behavior* in and out of equilibrium is essential in understanding perfect equilibria). The workhorse solution concept is thus *Bayesian perfect equilibrium*, which combines Harsanyi's (1967–68) solution concept for games of incomplete information with the credibility constraints of perfect equilibrium. See also Kreps and Wilson (1982a) for an articulation of the slightly more sophisticated sequential equilibrium solution concept, and Fudenberg and Tirole, Chapter 5 in this Handbook, for a further

[115] This result need not carry over to games in which the investment decisions can be modified quickly (the limiting case of this being the standard pricing supergame with constant marginal costs). Benoit and Krishna also study a model with this type of "flexible capacity".

[116] The topsy-turvy principle (large capacities allow firms to behave very competitively and hence support collusion) has now infected the investment game, since the continuation game following capacity choices is a supergame.

discussion of refinements of solution concepts in dynamic games with incomplete information.

What sort of uncertain variables are important for oligopolistic rivalry, and how might a firm be able to manipulate its rivals' expectations about these variables? The natural candidates are cost parameters and demand parameters. In quantity competition, for example, firm 1 benefits if firm 2 believes either (a) that firm 1's costs are low, (b) that firm 2's own costs are high, or (c) that demand is low.[117] Therefore, firm 1 will try to take actions to convince firm 2 that firm 1's costs are low. The theory of signaling tells us that for an action to signal low costs, it must be easier to undertake by a firm whose costs really are low. The obvious strategy is to expand output as a signal of lower costs. In equilibrium, today's output by firm 1 affects firm 2's beliefs about firm 1's costs. Of course, firm 2 understands firm 1's signaling incentives, and in a separating equilibrium it infers firm 1's costs perfectly. Therefore, firm 1's "disinformation" campaign is not successful; but the campaign nonetheless has an effect on firm 1's behavior. I refer the reader to Wilson (1983) and especially to Roberts (1987) for a more extensive discussion of these incomplete information models than is provided here.

Much of the literature on information manipulation has focused on the possibility of entry deterrence. This emphasis is natural, since it permits one to examine models with a single incumbent engaging in strategic signaling. The usual reference here is Milgrom and Roberts (1982a). Again, see Gilbert, Chapter 8 in this Handbook. Other papers study information manipulation, or "bluffing", as a predatory tactic; see Kreps and Wilson (1982b), Milgrom and Roberts (1982b), or Salop and Shapiro (1980). These applications of strategic signaling are covered in Chapter 9 on predation by Ordover and Saloner (1987) in this Handbook. Kreps, Milgrom, Roberts and Wilson (1982) come much closer to oligopoly theory in applying these same ideas to the repeated prisoner's dilemma game.

While little work has been done on information models in the context of dynamic oligopoly, many of the principles discovered in the entry deterrence and predation literatures have implications for oligopolistic behavior in markets where private information is significant. In repeated Cournot models with private demand information, for example, we expect more competitive behavior as firms try to send unfavorable demand information to their rivals [see Riordon (1985) again]. With uncertain information about firm-specific costs, we would again expect more competitive behavior than repeated Cournot, as firms attempt to signal their low costs to their competitors. But if a quantity-setting firm is signaling *industrywide* cost conditions via its outputs, we may have less competi-

[117]See the related discussion in Section 4 above. And note that (a) and (c) are reversed in the case of pricing competition. As usual, strategic incentives depend upon the slope of the reaction curves.

tive behavior.[118] And these results would probably be reversed if firms set prices rather than quantities. The incomplete information theories also can be applied to study a declining industry, as the firms engage in a war of attrition [although this situation has only been analyzed for the case of uncertain fixed costs – see Fudenberg and Tirole (1986b)], with each trying to send a signal to the other that will induce exit.

5.6. Simple state–space games

Clearly, there are many more dimensions of industry behavior that might serve as state variables and be subject to strategic control. I provided many such examples in Section 4 above. Here I indicate (without any claim of being comprehensive) a very few simple state variables that have been or might be examined in infinite horizon models. I see this as a promising area for future research, constrained mainly by the tractability of differential games.

(1) *Sunk investments.* A simple state variable is the extent of commitment to the market made by various sellers in the form of sunk investments. Models of this sort can become tractable if the state–space and the reduced-form profit functions are kept very simple. Gilbert and Harris (1984) did this by restricting attention to unit lumps of capital, so the state variable was a non-negative integer (as well as the date, since they studied a growing market). Likewise, Dixit and Shapiro (1986) consider the number of firms currently committed to the market (where each firm can either be "IN" or "OUT"). By assuming that flow profits are a simple reduced-form function of the number of firms (or plants), the game can be solved.

(2) *Learning-by-doing or network externalities.* With learning-by-doing or network externalities, previous sales affect the firms' current positions. With learning-by-doing, these sales affect current and future costs; with network externalities, it is demand that depends upon past sales. If the learning spills over across firms so that each firm learns equally from its own production and from production by rivals, the relevant state variable is cumulative industrywide sales; see Stokey (1986) for a fine analysis of this case. Cumulative industrywide sales also constitute the state variable in the case of network externalities where all products are fully compatible (so that a consumer only cares how many other consumers in total have bought, not which brands they purchased). Without perfect spillovers in the case of learning-by-doing, or without perfect product compatibility in the networks case, the state variables are cumulative sales by each of the firms. See Spence (1981) and Farrell and Saloner (1986), respectively.

[118]This is the oligopoly theory parallel to Harrington's (1986) point that a firm would lower its output to deter entry if it is signaling industrywide costs, the reverse of Milgrom and Roberts (1982a) finding in the case of firm-specific costs.

(3) *Consumer switching costs.* With consumer lock-in, previous choices by consumers affect current and future rivalry. Now the simple state variables are the market shares, say, of the rival firms. Farrell and Shapiro (1987) are able to solve a simple game of this sort by assuming identical tastes and identical switching costs across consumers, thereby greatly narrowing down the state space.

(4) *Shallow pockets.* In many markets, firms find it increasingly difficult or costly to attract financing as they attempt to borrow more money. Given these borrowing constraints, the firms' profits at a given date affect their investment opportunities at later dates. Then the firms' cash reserves constitute state variables, and aggressive pricing by one firm may serve a strategic purpose of undermining its rival's financial standing. Judd and Petersen (1986) have explored this type of effect in the context of entry deterrence.

(5) *R&D competition.* In dynamic R&D competition, the natural state variables are the firms' current progress on their R&D projects. See especially Fudenberg, Gilbert, Stiglitz and Tirole (1983), Harris and Vickers (1985), and Grossman and Shapiro (1987) for simple R&D state–space games in which a firm may have a lead or fall behind in the race.

This list of state variables in dynamic games is by no means exhaustive; I have omitted, for example, games with exogenously changing conditions in which the only state variable is time itself. But this brief list should indicate the range of possibilities for dynamic, state–space models. Further work in these and related areas may prove fruitful, as virtually all of the work on these topics is confined to two-period models. The main barrier to further progress at this point in time is simply tractability.

6. Conclusions

Having warned the reader at the outset that there are many theories of oligopoly, I am left with the task of identifying the lessons learned from the collection of models discussed above. I would emphasize my view that the variety of models of oligopolistic interactions is a virtue, not a defect. I would seek a single theory of oligopolistic behavior no more than I would a single set of behavioral rules for the survival of all species. Yes, some types of behavior, such as protecting one's young or developing the ability to survive on a varied diet, are beneficial to a wide variety of species, but the principle that tacit collusion works better when defections can be swiftly detected and punished has comparable generality. And just as the tactic of strategic investment works only in certain industries, survival based on being large and powerful and thereby having the ability to fend off predators is utilized by only a modest number of species.

What then are the lessons we can draw from the various models surveyed here? I regard these game-theoretic models as providing the industry analyst with a bag of tools. In other words, we have been able to identify quite a large number of strategic considerations that come into play when there are large firms and enough sunk costs so that threats of entry are not the primary determinate of industry behavior, at least for the short to medium run. One class of strategic behavior revolves around firms' efforts to tacitly collude, i.e. to favorably solve their basic prisoner's dilemma problem. Here we have learned about the factors that tend to facilitate collusion, e.g. careful monitoring of rival actions, the ability to write contracts with automatic price matching provisions, or the ability to change prices or production levels quickly in response to other firms' actions. The second class of behavior that we understand better on the basis of these models is that of strategic investment, where investment ranges from expenditures on sunk physical capital to strategic manipulation of a rival's information about market conditions. While there are no general results in the area of strategic investment, we have enough examples to understand quite well the strategic role of many business practices. What we are most in need of now are further tests of the empirical validity of these various theories of strategic behavior.[119]

Let me close with a sort of user's guide to the many oligopoly models I have discussed. By "user", I mean one who is attempting to use these models to better understand a given industry (not someone out to build yet another model). Here is where the "bag of tools" analogy applies. After learning the basic facts about an industry, the analyst with a working understanding of oligopoly theory should be able to use these tools to identify the main strategic aspects present in that industry.[120] One industry may be competitive because rapid expansions in capacity are possible in short order and consumers are willing to switch suppliers in response to small price differentials. In another industry, advertising may serve a key strategic role, since brand loyalty is significant. Yet another industry may succeed in achieving a tacitly collusive outcome because secret price-cutting is impossible. And so on. Hopefully, as further progress is made, we will learn about additional modes of strategic behavior and understand more fully the strategies already identified.

References

Abreu, D. (1986) 'Extremal equilibria of oligopolistic supergames', *Journal of Economic Theory*, 39:191–225.

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

[119]See Chapter 17 by Bresnahan in this Handbook for a survey of the relevant empirical findings.

[120]Of course, entry conditions may be such that the interactions among active producers are of secondary importance. This depends very much on the magnitude of sunk costs in the industry. I refer the reader to the chapters on contestability and entry deterrence for a working understanding of these essential components of industry analysis.

Abreu, D., Milgrom, P. and Pearce, D. (1987) 'Information, timing, and repeated partnership', Harvard University, unpublished.

Abreu, D., Pearce, D. and Stacchetti, E. (1986) 'Optimal cartel equilibria with imperfect monitoring', *Journal of Economic Theory*, 39:251–269.

Anderson, R. (1985) 'Quick-response equilibrium', University of California, Berkeley, unpublished.

Arvan, L. (1985) 'Some examples of dynamic Cournot duopoly with inventory', *Rand Journal of Economics*, 16:569–578.

Axelrod, R. (1984) *The Evolution of Cooperation*. New York: Basic books.

Beckman, M. (1965) 'Edgeworth–Bertrand duopoly revisited', in: R. Henn, ed., *Operations Research Verfahren, III*. Verlag Anton Hain: Meisenhein.

Benoit, J.P. and Krishna, V. (1985) 'Finitely repeated games', *Econometrica*, 53:905–922.

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

Bergstrom, T. (1979) 'Cournot equilibrium with factor markets', University of Michigan.

Bergstrom, T. and Varian, H. (1985a) 'Two remarks on Cournot equilibria', *Economics Letters*, 19:5–8.

Bergstrom, T. and Varian, H. (1985b) 'When are Nash equilibria independent of the distribution of agents' characteristics?' *Review of Economic Studies*, 52:715–718.

Bernheim, D. and Ray, D. (1987) 'Collective dynamic consistency in repeated games', Stanford University, unpublished.

Bernheim, B.D. and Whinston, M.D. (1985) 'Common marketing agency as a device for facilitating collusion', *Rand Journal of Economics*, 16:269–281.

Bertrand, J. (1883) 'Book review of *Theorie mathematique de la richesse sociale* and of *Recherches sur les principles mathematiques de la theorie des richesses*', *Journal des Savants*, 67:499–508.

Bowley, A. (1924) *Mathematical Foundations of Economics*. New York: Oxford University Press.

Brander, J.A. and Eaton, J. (1984) 'Product line rivalry', *American Economic Review*, 74:323–334.

Brander, J.A. and Lewis, T.R. (1985) 'Bankruptcy costs and the theory of oligopoly', University of British Columbia.

Brander, J.A. and Lewis, T.R. (1986) 'Oligopoly and financial structure: the limited liability effect', *American Economic Review*, 76:956–970.

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

Brander, J. and Spencer, B. (1985) 'Export subsidies and international market share rivalry', *Journal of International Economics*, 18:83–100.

Bresnahan, T.F. (1981) 'Duopoly models with consistent conjectures', *American Economic Review*, 71:934–94.

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

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

Cheng, L. (1985) 'Comparing Bertrand and Cournot equilibria: A geometric approach', *Rand Journal of Economics*, 16:146–152.

Cooper, T.E. (1986) 'Most-favored-customer pricing and tacit collusion', *Rand Journal of Economics*, 17:377–388.

Cournot, A.A. (1838) *Researches into the mathematical principles of the theory of wealth*, English edition of *Researches sur les principles mathematiques de la theorie des richesses*. New York: Kelley.

Cyert, R.M. and Degroot, M. (1970) 'Multiperiod decision models with alternating choice as a solution to the duopoly problem', *Quarterly Journal of Economics*, 84:410–429.

Dansby, R.E. and Willig, R.D. (1979) 'Industry performance gradient indices', *American Economic Review*, 69:249–260.

Dasgupta, P. and Maskin, E. (1986) 'The existence of equilibria in discontinuous economic games II: Applications', *Review of Economic Studies*, 53:27–41.

Daughety, A.F. (1985) 'Reconsidering Cournot: The Cournot equilibrium is consistent', *Rand Journal of Economics*, 16:368–379.

Davidson, C. and Deneckere, R. (1985a) 'Incentives to form coalitions with Bertrand competition', *Rand Journal of Economics*, 16:473–486.

Davidson, C. and Deneckere, R. (1985b) 'Excess capacity and collusion', Michigan State University, unpublished.

Davidson, C. and Deneckere, R. (1986) 'Long-run competition in capacity, short-run competition in price, and the Cournot model', *Rand Journal of Economics*, 17:404–415.

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

Dixit, A.K. (1982) 'Recent developments in oligopoly theory', *American Economic Review*, 72:12–17.

Dixit, A.K. (1986) 'Comparative statics for oligopoly', *International Economic Review*, 27:107–122.

Dixit, A.K. and Grossman, G. (1986) 'Targeted export promotion with several oligopolistic industries', *Journal of International Economics*, 20, forthcoming.

Dixit, A.K. and Kyle, A.S. (1985) 'The use of protection and subsidies for entry promotion and deterence', *American Economic Review*, 75:139–152.

Dixit, A.K. and Shapiro, C. (1986) 'Entry dynamics with mixed strategies', in: L.G. Thomas, ed., *Strategic planning*. Lexington, Mass.: Lexington Books, 63–79.

Dixon, H. (1986) 'The Cournot and Bertrand outcomes as equilibria in a strategic metagame', *Economic Journal*, 96:59–70.

Domowitz, I., Hubbard, R.G. and Petersen, B.C. (1986) 'Business cycles and the relationship between concentration and price–cost margins', *Rand Journal of Economics*, 17:1–17.

Doyle, C. (1985) 'Price matching strategies and oligopoly', Department of Applied Economics, University of Cambridge, unpublished.

Eaton, B.C. and Lipsey, R.G. (1980) 'Exit barriers are entry barriers: The durability of capital as a barrier to entry', *Bell Journal of Economics*, 11:721–729.

Eaton, B.C. and Lipsey, R.G. (1981) 'Capital, commitment, and entry equilibrium', *Bell Journal of Economics*, 12:593–604.

Eaton, J. and Grossman, G. (1986) 'Optimal trade and industrial policy under oligopoly', *Quarterly Journal of Economics*, 101:383–406.

Edgeworth, F. (1925) 'The pure theory of monopoly', *Papers Relating to Political Economy*, 1:111–142.

Farrell, J. and Maskin, E. (1987) 'Notes on renegotiation in repeated games', Harvard University, unpublished.

Farrell, J. and Saloner, G. (1986) 'Installed base and compatibility: Innovation, product preannouncements, and predation', *American Economic Review*, 76:940–955.

Farrell, J. and Shapiro, C. (1987) 'Dynamic competition with switching costs', *Rand Journal of Economics*, 19:123–137.

Farrell, J. and Shapiro, C. (1988) 'Horizontal mergers: An equilibrium analysis', John M. Olin Discussion Paper no. 17, Princeton University.

Fershtman, C. and Judd, K. (1984) 'Equilibrium incentives in oligopoly', *American Economic Review*, 77:927–940.

Fisher, I. (1898) 'Cournot and mathematical economics', *Quarterly Journal of Economics*, 12:119–138.

Friedman, J.W. (1968) 'Reaction functions and the theory of duopoly', *Review of Economic Studies*, 35:257–272.

Friedman, J.W. (1971) 'A noncooperative equilibrium for supergames', *Review of Economic Studies*, 38:1–12.

Friedman, J.W. (1977) *Oligopoly and the theory of games*. New York: North-Holland.

Friedman, J.W. (1983) *Oligopoly theory*. New York: Cambridge University Press.

Friedman, J.W. (1985) 'Cooperative equilibria in finite horizon noncooperative supergames', *Journal of Economic Theory*, 35:290–398.

Fudenberg, D. and Maskin, E. (1986) 'The folk theorem in repeated games with discounting and incomplete information', *Econometrica*, 54:533–554.

Fudenberg, D. and Tirole, J. (1983a) 'Learning by doing and market performance', *Bell Journal of Economics*, 14:522–530.

Fudenberg, D. and Tirole, J. (1983b) 'Capital as commitment: Strategic investment to deter mobility', *Journal of Economic Theory*, 31:227–250.

Fudenberg, D. and Tirole, J. (1984) 'The fat-cat effect, the puppy-dog ploy and the lean and hungry look', *American Economic Review Papers and Proceedings*, 74:361–366.

Fudenberg, D. and Tirole, J. (1986a) 'Dynamic models of oligopoly', in: A. Jacquemin, ed., *Fundamentals of pure and applied economics*, vol. 3. New York: Harwood.

Fudenberg, D. and Tirole, J. (1986b) 'A theory of exit in duopoly', *Econometrica*, 54:943–960.

Fudenberg, D., Gilbert, R., Stiglitz, J. and Tirole, J. (1983) 'Preemption, leapfrogging, and competition in patent races', *European Economic Review*, 22:3–31.
Gelman, J. and Salop, S. (1983) 'Capacity limitation and coupon competition', *Bell Journal of Economics*, 14:315–325.
Gertner, R. (1985) 'Dynamic duopoly with price inertia', Department of Economics, M.I.T., unpublished.
Ghemawat, P. and Nalebuff, B. (1985) 'Exit', *Rand Journal of Economics*, 16:184–193.
Gilbert, R. and Harris, R. (1984) 'Competition with lumpy investment', *Rand Journal of Economics*, 15:197–212.
Green, E. and Porter, R. (1984) 'Noncooperative collusion under imperfect price information', *Econometrica*, 52:87–100.
Grossman, G. and Shapiro, C. (1987) 'Dynamic R&D competition', *Economic Journal*, 97:372–387.
Grossman, S. (1981) 'Nash equilibrium and the industrial organization of markets with large fixed costs', *Econometrica*, 49:1149–1172.
Hahn, F. (1962) 'The stability of the Cournot oligopoly solution concept', *Review of Economic Studies*, 29:329–331.
Hall, R.L. and Hitch, C.J. (1939) 'Price theory and business behavior', *Oxford Economic Papers*, 2:12–45.
Harrington, J. (1986) 'Limit pricing when the potential entrant is uncertain of its cost function', *Econometrica*, 54:429–437.
Harris, C. and Vickers, J. (1985) 'Perfect equilibrium in a model of a race', *Review of Economic Studies*, 52:193–209.
Harsanyi, J.C. (1967–68) 'Games with incomplete information played by Bayesian players', Parts I, II, and III, *Management Science*, 14:159–182, 320–334, 486–502.
Holt, C. and Scheffman, D. (1986) 'Facilitating practices: The effects of advance notice and best-price policies', *Rand Journal of Economics*, 18:187–197.
Judd, K. (1985) 'Credible spatial preemption', *Rand Journal of Economics*, 16:153–166.
Judd, K. and Petersen, B. (1986) 'Dynamic limit pricing and internal finance', *Journal of Economic theory*, 39:368–399.
Katz, M. (1987) 'Game-playing agents: Contracts as precommitments', Department of Economics, Princeton University, unpublished.
Katz, M.L. and Shapiro, C. (1985) 'On the licensing of innovations', *Rand Journal of Economics*, 16:504–520.
Katz, M.L. and Shapiro, C. (1986a) 'Technology adoption in the presence of network externalities', *Journal of Political Economy*, 94:822–841.
Katz, M.L. and Shapiro, C. (1986b) 'Product compatibility choice in a market with technological progress', *Oxford Economic Papers*, 38:146–165.
Klemperer, P. (1985) 'Markets with consumer switching costs', *Quarterly Journal of Economics*, 102:375–394.
Klemperer, P. and Meyer, M. (1985) 'Price competition vs. quantity competition: The role of uncertainty', *Rand Journal of Economics*, 17:618–638.
Klemperer, P. and Meyer, M. (1986) 'Supply function equilibria under uncertainty', Graduate School of Business, Stanford University.
Kreps, D. and Scheinkman, J. (1983) 'Quantity pre-commitment and Bertrand competition yield Cournot outcomes', *Bell Journal of Economics*, 14:326–337.
Kreps, D. and Spence, A.M. (1985) 'Modelling the role of history in industrial organization and competition', in: G.R. Feiwel, ed., *Issues in contemporary microeconomics and welfare*. London: MacMillan, 340–378.
Kreps, D. and Wilson, R. (1982a) 'Sequential equilibrium', *Econometrica*, 50:863–894.
Kreps, D. and Wilson, R. (1982b) 'Reputation and imperfect information', *Journal of Economic Theory*, 27:253–279.
Kreps, D., Milgrom, P., Roberts, J. and Wilson, R. (1982) 'Rational cooperation in the finitely repeated prisoners' dilemma', *Journal of Economic Theory*, 27:245–252.
Krugman, P. (1984) 'Import protection as export promotion: International competition in the presence of oligopoly and economies of scale', in: H. Kierzkowski, *Monopolistic competition and international trade*. New York: Oxford University Press, 180–193.

Laitner, J. (1980) 'Rational duopoly equilibrium', *Quarterly Journal of Economics*, 95:641–662.

Levitan, R. and Shubik, M. (1972) 'Price duopoly and capacity constraints', *International Economic Review*, 14:326–337.

Mailath, G. (1985) 'Welfare in a simultaneous signaling duopoly model', Caress working paper no. 85-29, University of Pennsylvania.

Makowski, L. (1983) '"Rational conjectures" aren't rational, "reasonable conjectures" aren't reasonable', economic theory discussion paper no. 66, University of Cambridge.

Marshak, T. and Selten, R. (1978) 'Restabilizing responses, inertia supergames and oligopolistic equilibria', *Quarterly Journal of Economics*, 92:71–93.

Maskin, E. (1986) 'The existence of equilibrium with price-setting firms', *American Economic Review*, 76:382–386.

Maskin, E. and Tirole, J. (1987) 'A theory of dynamic oligopoly III: Cournot competition', *European Economic Review*, 31:947–968.

Maskin, E. and Tirole, J. (1988a) 'A theory of dynamic oligopoly I: Overview and quantity competition with large fixed costs', *Econometrica*, 56:549–569.

Maskin, E. and Tirole, J. (1988b) 'A theory of dynamic oligopoly II: Price competition, kinked demand curves, and Edgeworth cycles', *Econometrica*, 56:571–599.

Milgrom, P. and Roberts, J. (1982a) 'Limit pricing and entry under incomplete information: An equilibrium analysis', *Econometrica*, 50:443–459.

Milgrom, P. and Roberts, J. (1982b) 'Predation reputation and entry deterrence', *Journal of Economic Theory*, 27:280–312.

Milgrom, P. and Roberts, J. (1987) 'Information asymmetries, strategic behavior, and industrial organization', *American Economic Review Papers and Proceedings*, 77:184–193.

Nelson, R. and Winter, S. (1982) *An evolutionary theory of economic change*. Cambridge, Mass.: Harvard University Press.

Novshek, W. (1980) 'Cournot equilibrium with free entry', *Review of Economic Studies*, 47:473–486.

Novshek, W. (1985) 'On the existence of Cournot equilibrium', *Review of Economic Studies*, 52:85–98.

Osborne, M. and Pitchik, C. (1984) 'Price competition in a capacity-constrained duopoly', discussion paper no. 185, Department of Economics, Columbia University.

Pearce, D. (1987) 'Renegotiation-proof equilibria: Collective rationality and intertemporal cooperation', Yale University, unpublished.

Perry, M. (1982) 'Oligopoly and consistent conjectural variations', *Bell Journal of Economics*, 13:197–205.

Perry, M. and Porter, R. (1985) 'Oligopoly and the incentive for horizontal merger', *American Economic Review*, 75:219–227.

Porter, R. (1983a) 'Optimal cartel trigger-price strategies', *Journal of Economic Theory*, 29:313–338.

Porter, R. (1983b) 'A study of cartel stability: The joint executive committee, 1880–1886', *Rand Journal of Economics*, 14:301–314.

Prescott, E. and Visscher, M. (1977) 'Sequential location among firms with foresight', *Bell Journal of Economics*, 8:378–394.

Radner, R. (1980) 'Collusive behavior in noncooperative epsilon-equilibria of oligopolies with long but finite lives', *Journal of Economic Theory*, 22:136–154.

Radner, R. (1986) 'Repeated partnership games with imperfect monitoring and no discounting', *Review of Economic Studies*, 53:43–57.

Riordan, M. (1985) 'Imperfect information and dynamic conjectural variations', *Rand Journal of Economics*, 16:41–50.

Roberts, J. (1987) 'Battles for market share: Incomplete information, aggressive strategic pricing, and competitive dynamics', in: T. Bewley, ed., *Advances in economic theory, invited papers from the Vth World Congress*. New York: Cambridge University Press.

Rosenthal, R. (1981) 'Games of imperfect information predatory pricing and the chain-store paradox', *Journal of Economic Theory*, 25:92–100.

Rotemberg, J. and Saloner, G. (1985) 'Strategic inventories and the excess volatility of production', working paper no. 1650-85, Sloan School of Management, M.I.T.

Rotemberg, J. and Saloner, G. (1986a) 'A supergame-theoretic model of business cycles and price wars during booms', *American Economic Review*, 76:390–407.

Rotemberg, J. and Saloner, G. (1986b) 'Quotas and the stability of implicit collusion', N.B.E.R. working paper no. 1948.

Salant, S., Switzer, S. and Reynolds, R. (1983) 'Losses from horizontal merger: The effects of an exogenous change in industry structure on Cournot–Nash equilibrium', *Quarterly Journal of Economics*, 98:185–199.

Saloner, G. (1986) 'The role of obsolescence and inventory costs in providing commitment', Department of Economics, M.I.T.

Salop, S. (1986) 'Practices that credibly facilitate oligopoly co-ordination', in: J. Stiglitz and F. Mathewson, eds., *New developments in the analysis of market structure*. Cambridge: M.I.T. Press, 265–290.

Salop, S. and Scheffman, D. (1983) 'Raising rivals' costs', *American Economic Review Papers and Proceedings*, 73:267–271.

Salop, S. and Shapiro, C. (1980) 'Test market predation', Federal Trade Commission, unpublished.

Schelling, T. (1960) *The strategy of conflict*. New York: Oxford University Press.

Scherer, F.M. (1980) *Industrial market structure and economic performance*. Chicago: Rand McNally.

Schmalensee, R. (1983) 'Advertising and entry deterrence', *Journal of Political Economy*, 90:636–653.

Schumpeter, J. (1954) *History of economic analysis*. New York: Oxford University Press.

Seade, J. (1980) 'The stability of Cournot revisited', *Journal of Economic Theory*, 23;15–27.

Seade, J. (1985) 'Profitable cost increases and the shifting of taxation: Equilibrium responses of markets in oligopoly', Warwick University, unpublished.

Shaked, A. and Sutton, J. (1982) 'Relaxing price competition through product differentiation', *Review of Economic Studies*, 49:3–14.

Shapiro, C. (1986) 'Exchange of cost information in oligopoly', *Review of Economic Studies*, 53:433–446.

Shubik, M. (1959) *Strategy and market structure*. New York: Wiley.

Singh, N. and Vives, X. (1984) 'Price and quantity competition in a differentiated duopoly', *Rand Journal of Economics*, 15:546–554.

Sklivas, S. (1985) 'The strategic choice of managerial incentives', discussion paper no. 184, Center for the Study of Organizational Innovation, University of Pennsylvania.

Spence, A.M. (1976) 'Product selection, fixed costs, and monopolistic competition', *Review of Economic Studies*, 43:217–235.

Spence, A.M. (1977) 'Entry, capacity, investment and oligopolistic pricing', *Bell Journal of Economics*, 8:534–544.

Spence, A.M. (1979) 'Investment strategy and growth in a new market', *Bell Journal of Economics*, 10:1–19.

Spence, A.M. (1981) 'The learning curve and competition', *Bell Journal of Economics*, 12:49–70.

Spence, A.M. (1984) 'Cost-reduction, competition, and industry performance', *Econometrica*, 52:101–122.

Stanford, W. (1986) 'Subgame perfect reaction function equilibria in discounted duopoly supergames are trivial', *Journal of Economic theory*, 39:226–232.

Stigler, G. (1947) 'The kinky oligopoly demand curve and rigid prices', *Journal of Political Economy*, 55:432–449.

Stigler, G. (1964) 'A theory of oligopoly', *Journal of Political Economy*, 72:44–61.

Stokey, N. (1986) 'The dynamics of industrywide learning', in: W.P. Heller, R.M. Starr and D.A. Starrett, eds., *Equilibria analysis: Essays in honor of Kenneth J. Arrow, Volume II*. London: Cambridge University Press, 81–104.

Sweezy, P. (1939) 'Demand under conditions of oligopoly', *Journal of Political Economy*, 47:568–573.

Vives, X. (1984) 'Duopoly information equilibrium Cournot and Bertrand', *Journal of Economic Theory*, 34:71–94.

Vives, X. (1985) 'On the efficiency of Cournot and Bertrand competition with product differentiation', *Journal of Economic Theory*, 36(1):166–175.

Vives, X. (1986) 'Commitment, flexibility and market outcomes', *International Journal of Industrial Organization*, 4.

von Stackelberg, H. (1934) *Marktform und Gleichgewicht*. Vienna: Springer.

Wilson, R. (1983) 'Reputations in games and markets', Stanford University, unpublished.