11

Dynamic Games and First and Second Movers

In 2007, Apple introduced its first *iPhone*. While other advanced, wireless phones already existed (principally, RIM's *Blackberry*), the *iPhone* was the first to allow input through a multi-touch screen rather than with a stylus or keypad, and the first to be compatible with downloadable applications. Its music playing and web browsing capacities were also far superior to anything that then existed. In short, this was the first of the modern "smart" phones.

Competition was not long in coming. In 2008, Google released the operating system *Android* as an open-source platform available to other hardware and software producers. T-Mobile was first with its *HTC Dream* phone but Motorola and Samsung soon began marketing their own *Android*-based phones. As this text is being written, Apple still retains its dominant position (within the United States, at least) as the leading smartphone manufacturer. However, the *Android* operating system now powers more smartphones across all manufacturers combined.[1]

An essential feature of the above strategic interaction between Apple and its rivals is its sequential nature. That is, unlike the simultaneous games discussed in the previous two chapters, this rivalry evolved dynamically over time. First, Apple took an action, and then—after that action was taken and observed—its rivals, principally Google, responded with their own action. Such dynamic games are the focus of this chapter. In principal, these games can have many rounds of play, sometimes called stages. Here, we concentrate mostly on games with just two stages and, for convenience, just two firms. Typically, one firm will play in the first round, the first mover, and the other will play in the second round, the second mover.

Popular business literature is replete with stories about first movers.[2] Many of these focus on specific cases in which, like the Apple smartphone example, the first mover is an early entrant who gains a dominant position and subsequently faces competition from later entrants. In the late nineteenth century, for example, Campbell's emerged as the dominant maker of canned soup in North America. Similarly, Heinz was the first entrant in the UK canned-soup market. Later, Campbell's entered the UK market after Heinz, and similarly Heinz entered the US market after Campbell's. Yet the first mover in each market continues

[1] Nielsen Media Research, "Mobile Media Report: Q3 2011."
[2] Lieberman and Montgomery (1988).

to dominate. Campbell's has roughly 63 percent of the US market, but only 9 percent of the UK market, whereas Heinz has a 41 percent market share in the United Kingdom and a relatively minor market share in the United States.[3]

Sequential entry and dynamic games merit explicit economic analysis, partly because they are so common but also because, as the Campbell's and Heinz cases illustrate, they suggest important advantages for the first entrant into a market in facing competition from later entrants. Remember, entry is a key part of the competitive market's success story as an allocative mechanism. It is entry—or the threat of entry—that erodes the market power of an established firm and that transforms otherwise monopolized markets into competitive ones. This enforcement mechanism will, however, be weakened if there is any asymmetry between established firms and later rivals. Understanding such asymmetry is therefore an important issue in industrial economics.

In the next two chapters, we explore the entry process in detail when the firms—the entrants and the incumbents—are strategic players in the market place. At this juncture, the point to realize is that entry is a sequential process—some firms enter early and some enter late. Thus developing an understanding of dynamic games is good groundwork for our later investigation of entry and entry deterrence in oligopoly markets.

We first examine quantity and price competition when firms move sequentially rather than simultaneously. We will discover again that price and quantity competition are different, and depending on the kind of competition, there can be first-mover or second-mover advantages. This raises the interesting question of whether and how a firm can become either a first- or second-mover. Often the key to achieving the desired position and the associated higher profits is the ability of the firm to make a credible commitment to its strategy when the market opens for trade. We examine what credibility means in game theory and how it affects our equilibrium solution concept for dynamic models.

Simultaneous games, such as the traditional Cournot or Bertrand model, describe a once-and-for-all market interaction between the rival firms. In some sequential games as well there is only one market period where trade takes place, although this might occur in several stages. However, the more likely scenario is that rival firms interact and trade today in the market and then interact again in the future. Moreover, the competing firms understand the likelihood of future interactions today. Repeating the market interaction over and over again gives rise to a somewhat different type of dynamic game, usually called a repeated game. We defer our discussion of repeated games until Chapter 14.

11.1 THE STACKELBERG MODEL OF QUANTITY COMPETITION

The duopoly model of Stackelberg (1934) is similar to the Cournot model except for one critically important difference. The two firms now choose quantities *sequentially* rather than *simultaneously*. The firm that moves first and chooses its output level first is the leader firm. The firm that moves second is the follower firm. The sequential choice of output is what makes the game dynamic. However, the firms trade their goods on the market only once and their interaction yields a "once-and-for-all" market-clearing outcome.

Let market demand again be represented by a linear inverse demand function $P = A - BQ$. Firm 1 is the leader who moves first and firm 2 is the follower who chooses its output *after* the choice of the leader is made. Each firm has the same constant unit

[3] See, for example, Sutton (1991).

cost of production c. Total industry output Q equals the sum of the outputs of each firm, $Q = q_1 + q_2$.

Firm 1 acts first and chooses q_1. How should it make this choice? Both firms are rational and strategic and both firms know this, and know that each other knows this. As a result, firm 1 will make its choice taking into account its best guess as to firm 2's rational response to its choice of q_1. In other words, firm 1 will work out firm 2's best response to each value of q_1, incorporate that best response into its own decision-making, and then choose the q_1 which, given firm 2's best response, maximizes firm 1's profit.

We can solve for firm 2's best response function q_2^* exactly as we did in the Cournot model in Chapter 9. For any choice of output q_1, firm 2 faces the inverse demand and marginal revenue curves:

$$\begin{aligned} P &= (A - Bq_1) - Bq_2 \\ MR_2 &= (A - Bq_1) - 2Bq_2 \end{aligned} \tag{11.1}$$

Setting marginal revenue equal to marginal cost yields firm 2's best response q_2^* as the solution to the first-order condition:

$$A - Bq_1 - 2Bq_2^* = c \tag{11.2}$$

from which we obtain:

$$q_2^* = \frac{(A - c)}{2B} - \frac{q_1}{2} \tag{11.3}$$

If firm 1 is rational, firm 1will understand that equation (11.3) describes what firm 2 will do in response to each possible choice of q_1. We can summarize equation (11.3) by $q_2^*(q_1)$. Anticipating this behavior by firm 2, firm 1 can substitute $q_2^*(q_1)$ for q_2 in its demand function so that its inverse demand function may be written as:

$$P = A - Bq_2^*(q_1) - Bq_1 = \frac{A + c}{2} - \frac{B}{2}q_1, \tag{11.4}$$

In turn, this implies that its profit function is:

$$\Pi_1(q_1, q_2^*(q_1)) = \left(\frac{A + c}{2} - \frac{B}{2}q_1 - c\right)q_1 = \left(\frac{A - c}{2} - \frac{B}{2}q_1\right)q_1 \tag{11.5}$$

Note that this substitution results in firm 1's demand and profits being dependent only on its own output choice, q_1. This is because firm 1 effectively sets q_2 as well, by virtue of the fact that q_2 is chosen by firm 2 in response to q_1 according to firm 2's best response function, *and firm 1 anticipates this*. In other words, the first-mover correctly predicts the second-mover's best response and incorporates this prediction into its decision-making calculus.

To solve for firm 1's profit-maximizing output q_1^* we find the marginal revenue curve associated with firm 1's demand curve in (11.4), that is, $MR_1 = \dfrac{A + c}{2} - Bq_1$, and find the output q_1^* at which marginal revenue is equal to marginal cost. Alternatively, we could maximize the profit function of equation (11.5) with respect to q_1^*. Either way we find that:

$$q_1^* = \frac{(A-c)}{2B} \tag{11.6}$$

Given this output choice by firm 1, firm 2 selects its best response as given by equation (11.3), which yields:

$$q_2^* = \frac{(A-c)}{4B} \tag{11.7}$$

Together, equations (11.6) and (11.7) describe the Stackelberg-Nash equilibrium production levels of each firm. Note that the leader's output is exactly equal to the level of output chosen by a simple uniform-pricing monopolist. This is a well-known feature of the Stackelberg model when demand is linear and costs are constant.

The total industry production is of course the sum of the two outputs shown in equations (11.6) and (11.7). This sum is: $Q^S = \frac{3(A-c)}{4B}$. Compare this market output with the earlier Cournot-Nash equilibrium industry output $Q^C = \frac{2(A-c)}{3B}$. Clearly, the Stackelberg model yields a greater industry output. Accordingly, the equilibrium price is lower in the Stackelberg model than it is in the Cournot model. The price and output results are illustrated in Figure 11.1.

A central feature of the Stackelberg model is the difference in the relative outcome of the two firms. Recall that from the standpoint of both consumer preferences and production techniques, the firms are identical. They produce identical goods and do so at the same constant unit cost. Yet, because one firm moves first, the outcome for the two firms is different. Comparing q_1^* and q_2^* reveals that the leader gets a far larger market share and earns a much larger profit than does the follower. Moving first clearly has advantages. Alternatively, entering the market late has its disadvantages.

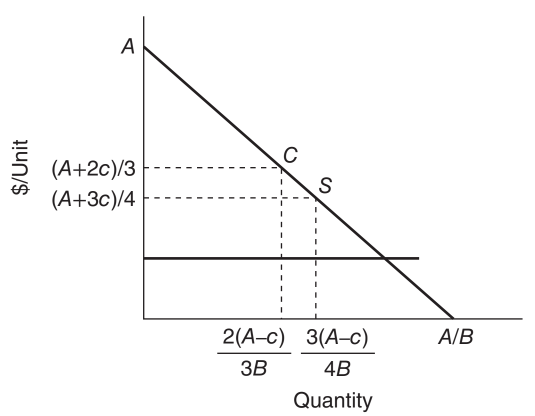

Figure 11.1 The Cournot and Stackelberg outcomes compared
C = Cournot Equilibrium; S = Stackelberg Equilibrium

An interesting additional aspect of the disadvantaged outcome for firm 2 in the Stackelberg model is that this outcome occurs even though firm 2 has full information regarding the output choice of q_1. Indeed, firm 2 actually observes that choice before selecting q_2. In the Cournot duopoly model, firm 2 did not have such concrete information. Because the Cournot model is based upon simultaneous moves, each firm could only make a (rational) guess as to its rival's output choice. Paradoxically, firm 2 does worse when it has complete information about firm 1's choice (the Stackelberg case) than it does when its information is less than perfect (the Cournot case). This is because saying the information is concrete amounts to saying that firm 1's choice—at the time that firm 2 observes it—is irreversible. In the Stackelberg model, by the time firm 2 moves, firm 1 is already fully committed to $q_1 = \frac{(A-c)}{2B}$. In the Cournot context, $q_1 = \frac{(A-c)}{2B}$ is *not* a best response to the choice $q_2 = \frac{(A-c)}{4B}$ and so firm 2 would not anticipate that firm 1 would produce that quantity. In contrast, in the Stackelberg model we do not derive firm 1's choice as a best response to $q_2 = \frac{(A-c)}{4B}$. Instead, we derived firm 1's output choice as the profit-maximizing output when firm 1 correctly anticipates that firm 2's decision is to choose its best value of q_2 *conditional upon* the output choice already made by firm 1. It is this fact that reflects the underlying assumption of sequential moves that distinguishes the Stackelberg model.

Stackelberg's modification to the basic Cournot model is important. It is a useful way to capture the observed phenomenon that one firm often has a dominant or leadership position in a market. The Stackelberg model reveals that moving first can have its advantage and therefore can be an important aspect of strategic interaction.

11.1

Practice Problem

Consider the following game. Firm 1, the leader, selects an output q_1, after which firm 2, the follower, observes the choice of q_1 and then selects its own output q_2. The resulting price is one satisfying the industry demand curve $P = 200 - q_1 - q_2$. Both firms have zero fixed costs and a constant marginal cost of 60.

a. Derive the equation for the follower firm's best response function. Draw this equation on a graph with q_2 on the vertical axis and q_1 on the horizontal axis. Indicate the vertical intercept, horizontal intercept, and slope of the best response function.
b. Determine the equilibrium output of each firm in the leader-follower game. Show that this equilibrium lies on firm 2's best response function. What are firm 1's profits in the equilibrium?
c. Now let the two firms choose their outputs simultaneously. Compute the Cournot equilibrium outputs and industry price. Who loses and who gains when the firms play a Cournot game instead of the Stackelberg one?

11.2 SEQUENTIAL PRICE COMPETITION

Now consider what happens if we keep the dynamic game framework above but change the strategic variable from output to price. Otherwise, the model is exactly the same as before. Each firm produces an identical good at the same, constant marginal cost, c, and consumers will purchase the good from the lowest-priced firm. If they set the same prices, then each firm will serve half the market.

In setting its price, firm 1 must of course anticipate firm 2's best response. Clearly firm 2 will have an incentive to price slightly below firm 1's price whenever firm 1 sets a price greater than unit cost c and less than or equal to the monopoly price. In that case, by undercutting, firm 2 will serve the entire market and earn all the potential profits. On the other hand, if firm 1 sets a price less than unit cost c, then firm 2 will not match or undercut firm 1's price because firm 2 has no interest in making any sales when each unit sold loses money. Finally, if firm 1sets a price equal to unit cost c, firm 2's best response is to match it. The anticipated behavior of firm 2 in stage 2 puts firm 1 in a tight bind. Any price greater than unit cost c results in zero sales and there is no sense in setting a price less than c. The best firm 1 can do then is set a price equal to unit cost c. Firm 2's best response in the next stage is to match firm 1's price.

As we saw in Chapter 10, however, matters are very different if the two firms are not selling identical products. In this case, not all consumers buy from the lower-priced firm. Product differentiation changes the outcome of price competition. To illustrate the nature of price competition with differentiated products, recall the spatial model of product differentiation that we developed previously. There is a product spectrum of unit length along which consumers are uniformly distributed. Two firms supply this market. One firm has the address or product design $x = 0$ on the line whereas the other has location $x = 1$. Each of the firms has the same constant unit cost of production c.

A consumer's location in this market refers to the consumer's most preferred product or style. "Consumer x" is located distance x from the left-hand end of the market. Consumers differ regarding which variant or location of the good they consider to be the best, or their ideal product, but are identical in their reservation price V for their most preferred product. We assume that the reservation price V is substantially greater than the unit cost of production c. Each consumer buys at most one unit of the product. If consumer x purchases a good that is not the ideal product, a utility loss of tx is incurred if good 1 (located at $x = 0$) is consumed; a utility loss of $t(1 - x)$ is incurred if good 2 (located at $x = 1$) is incurred.

The two firms compete for customers by setting prices, p_1 and p_2, respectively. However, unlike the simple Bertrand model, firm 1 sets its price p_1 first, and then firm 2 follows by setting p_2. In order to find the demand facing the firms at prices p_1 and p_2 we proceed as in the previous chapter by identifying the marginal consumer x^m, who is indifferent between buying from either firm 1 or firm 2. Indifference means that consumer x^m gets the same consumer surplus from either product and so satisfies the condition:

$$V - p_1 - tx^m = V - p_2 - t(1 - x^m) \tag{11.8}$$

From equation (11.8) we find that the address of the marginal consumer x^m is:

$$x^m\,(p_1, p_2) = \frac{(p_2 - p_1 + t)}{2t} \tag{11.9}$$

At any set of prices, p_1 and p_2, all consumers to the left of x^m buy from firm 1, and all those to the right of x^m buy from firm 2. In other words, x^m is the fraction of the market buying from firm 1 and $(1 - x^m)$ is the fraction buying from firm 2. If the total number of consumers is denoted by N and they are uniformly distributed over the product spectrum, the demand function facing firm 1 at any price combination (p_1, p_2) is:

$$D^1(p_1, p_2) = x^m(p_1, p_2) = \frac{(p_2 - p_1 + t)}{2t}N \tag{11.10}$$

Similarly, firm 2's demand function is:

$$D^2 = (p_1, p_2) = (1 - x^m(p_1, p_2)) = \frac{(p_1 - p_2 + t)}{2t}N \tag{11.11}$$

Firm 1 acts first and sets its price p_1. In doing so, firm 1 anticipates firm 2's best response to the price p_1 that firm 1 sets. In other words, firm 1 works out firm 2's best response to each possible price p_1, and then chooses its profit-maximizing price p_1 given firm 2's best response to that price. We can solve for firm 2's best response function p_2^* exactly as we did in Section 3 of Chapter 10 [equation (10.14)]. It is

$$p_2^* = \frac{p_1 + c + t}{2} \tag{11.12}$$

Firm 1 knows that equation (11.12) describes what firm 2 will do in response to each price p_1 that firm 1 could set. We can summarize equation (11.12) by $p_2^*(p_1)$. Firm 1 knows that if it sets first a price p_1, then firm 2 will set a price $p_2^*(p_1)$. As a result, firm 1's demand (11.10) becomes:

$$D^1(p_1, p_2^*(p_1)) = q_1 = \frac{(p_2^*(p_1) - p_1 + t)}{2t}N = \frac{N}{4t}(c + 3t - p_1) \tag{11.13}$$

Firm 1's inverse demand curve is therefore:

$$p_1 = c + 3t - \frac{4t}{N}q_1 \tag{11.14}$$

By the twice-as-steep rule, firm 1's marginal revenue is given by: $MR = c + 3t - \frac{8t}{N}q_1$. Equating this with marginal cost c then yields firm 1's optimal output as $q_1^* = \frac{3N}{8}$. In turn, this implies that firm 1's profit-maximizing price is:

$$p_1^* = c + \frac{3t}{2} \tag{11.15}$$

Given this choice of price by firm 1, firm 2 uses equation (11.12) to choose its best price:

$$p_2^* = c + \frac{5t}{4} \tag{11.16}$$

Recall that in Section 10.3, when we analyzed this market with simultaneous price competition the result was the symmetric outcome in which each firm set the same prices $p_1^* = p_2^* = c + t$, and each served half the market. Things change, however, in the sequential game. Both firms now charge higher prices but the prices are no longer equal. The price leader, firm 1, now sets a higher price than the follower, firm 2. As a result, firm 2 now serves the larger market share—5/8 relative to the 3/8 served by firm 1.

Note that firm 1's price leadership has helped both firms. In the simultaneous game, each earned a profit of $Nt/2$. Now, firm 1 earns a profit equal to $9Nt/16$, while firm 2 earns an even greater profit equal to $25Nt/32$. That is, while both firms are better off under the price leadership game (and consumers are worse off), it is the second or follower firm that gains the most. There is then an important first-mover *dis*advantage in this price game. Once firm 1 commits to a price, it becomes a fixed target for firm 2 to undercut. Yet because goods are differentiated it is difficult for firm 2 to steal all of firm 1's customers unless it prices really low. Knowing that this is not profitable and that firm 2 will therefore not undercut its price too much is what leads firm 1 to set a fairly high—and more profitable—price to begin with. If consumers regarded the goods as perfect substitutes so that the parameter $t = 0$, prices would again fall to marginal cost as in the original Bertrand model, and there would be no first- or second-mover advantage.

11.2 **Practice Problem**

Let there be two hair salons located on Main Street, which is one mile long. One is located at the east end of town, $x = 0$, and the other is located at the east end, $x = 1$. There are 100 potential customers who live along the mile stretch, and they are uniformly spread out along the mile. Consumers are willing to pay \$50 for a haircut done at their home. If a consumer has to travel there and back to get a haircut then a travel cost of \$5 per mile is incurred. Each salon has the same unit cost equal to \$10 per haircut.

a. Suppose the East End Salon posts its price for a haircut first and then the West End Salon posts its price for a haircut. What prices will the two salons set? How many customers does each salon serve? What are the profits?
b. Compare the prices to the ones we found when the two salons set their prices simultaneously [Chapter 10, equation (10.15)]. Explain why prices changed in the way they did.

Firms generally seem to do better when they compete sequentially in prices than when they compete sequentially in output. The average price is higher and both firms earn higher profits when price competition is sequential rather than simultaneous. In contrast, the industry price falls and only one firm earns higher profit when quantity competition becomes sequential rather than simultaneous. This difference is related to another distinction: whereas it is the first mover who has the clear advantage in the quantity game, in the price game, it is the firm who moves last that does best.[4]

It is important to understand the central reason that the dynamic games above yield such different outcomes from their simultaneous counterparts. Again, the reason is that sequential play means that when the follower moves, firm 1's move is taken as irreversible or given. To put it differently, the game only becomes truly sequential if firm 1 can make a *credible commitment* not to alter its choice after firm 2 has played. If for example, firm 1 could alter its output or price after firm 2 moves, the Nash equilibrium in either the output or price games discussed above would be the same as in the simultaneous games of Chapters 9 and 10. Because credibility is so important, we should expect that the firms playing dynamic games will also distinguish between credible strategies and non-credible ones. So, we need to understand what makes strategies credible in dynamic games.

We explore dynamic credibility in the next section. We do so in the context of a market entry game. This is a game that has been of great interest to industrial organization economists. In this game, the incumbent firm announces a strategy promising its reaction to actual entry. The entrant then moves by deciding whether or not to enter, taking account the incumbent's threatened reaction. The question is what threats are credible.

[4] This would not be the case if the market were instead vertically differentiated and the first mover was also the firm with the highest quality good.

Reality Checkpoint

First-Mover Advantage in the TV Market: More Dishes and Higher Prices

When a firm markets a new good or service, consumers understand that it may take time to learn how to use it in such a way that one gets full use of all the product's features. For example, it takes experience to learn the full use of an Apple *iPhone* or how to trade on e-Bay. Gabszewicz, Pepall, and Thisse (GPT) (1992) develop a two-stage model to show how consumer learning may confer a first-mover advantage to the first firm to market a new product. Firm 1 leads with its version of a new good. Firm 2 enters in stage two with its own variant of the same good. GPT argue that those consumers who bought firm 1's product in stage 1 will know now how to use it effectively but they will not know that for firm 2's product. As a result, they will tend to prefer firm 1's good even if firm 2 sells at a lower price.

GPT show that the pricing implications can be quite novel. Foreseeing the advantage that learning confers against a later entrant, firm 1 has an incentive to price very low in the first stage so as to induce a lot of consumers to try and to become experienced with its product before firm 2 enters. This creates a large group of captive consumers willing to pay a higher price for firm 1's product in stage 2 now that they know how the product works. Thus when firm 2 enters, firm 1 actually raises its price and still retains a large number of consumers who cannot be bothered to learn how to use the entrant's different version of the good. Not only is there a first-mover advantage but prices rise at the very time that new competition emerges—the opposite of what simple textbook analysis often implies.

The rivalry between cable and dish TV may illustrate this point. Cable was generally first in this market and spread fast with over 70 percent of American homes now receiving cable service. Direct broadcast satellite (DBS) TV that consumers receive through a satellite dish followed. Textbook analysis would suggest that DBS competition would lead to lower cable prices. However, Goolsbee and Petrin (2004) found that, to the contrary, the spread of dish TV led to an increase in the annual cable fee of about \$34.68, precisely the first-mover advantage noted by Gabszewicz, Pepall, and Thisse.

Source: Gabszewicz, J., Pepall, L. and J-F. Thisse, 1992. "Sequential Entry with Brand Loyalty Caused by Consumer Learning-By-Doing," *Journal of Industrial Economics* 60 (December): 397–416; and Goolsbee, A., and A. Petrin, 2004. "The Consumer Gains from Direct Broadcast Satellite and Competition with Cable TV," *Econometrica* 72 (March): 351–81.

11.3 CREDIBILITY OF THREATS AND NASH EQUILIBRIA FOR DYNAMIC GAMES

We begin by introducing a concept that is critical to all dynamic games, namely, that of a subgame. A subgame is a part of an entire game that can stand alone as a game in itself. A proper subgame is a game within a game. Simultaneous games cannot have subgames, but dynamic games can. An example of a subgame in a two period model is the competition in the second period, which is a one-shot game within the larger two-period game.

Closely related to the notion of subgame is the concept of subgame perfection, first introduced by Nobel Prize winner Reinhard Selten (1978). It is the concept of subgame perfection that permits us to understand whether a firm's strategy is credible in a dynamic game. The term sounds very technical but it is actually quite simple. Basically, subgame perfection means that if a strategy chosen at the start of a game is optimal, it must be optimal to stick with that strategy at every later juncture in the game as play progresses.

It is easier to understand the concept of subgame perfection by seeing its application in practice. Imagine then a dynamic game between two software firms, one a giant called Microhard who is the incumbent firm in the market and the other an upstart firm, Newvel, who wishes to enter the market. In this game the potential entrant, Newvel, chooses either to enter Microhard's market or to stay out. If Newvel stays out it earns a normal profit from being somewhere else in the economy, say $\Pi = 1$, and Microhard continues to earn a monopoly profit in the software market, say $\Pi = 5$. If Newvel enters the market then Microhard can choose either to accommodate the new entrant and share the market or to fight the new entrant by slashing prices. If Microhard accommodates Newvel's entry, then each firm earns a profit $\Pi = 2$. If, on the other hand, Microhard fights, then neither firm makes any profit so each firm earns $\Pi = 0$.

Dynamic games with moves in sequence require more care in presentation than single-period, simultaneous games. In a simultaneous game, a firm moves once and simultaneously and so its *action* is the same as its *strategy*. For a dynamic game, a firm's strategy is a complete set of instructions that tell the firm what *actions* to pick at every conceivable situation in the game. However, we can begin the analysis of this entry game between Microhard and Newvel using a payoff matrix of the type introduced in Chapter 9. For the game at hand, this matrix is shown in Table 11.1.

Start with the combination (Enter, Fight). This *cannot* correspond to an equilibrium. Enter will lead Newvel to come into the market. If Microhard has adopted the Fight strategy, it must respond to such an entry very aggressively. Yet, as the payoff matrix makes clear, such an aggressive action is not Microhard's best response to entry by Newvel. Now try (Enter, Accommodate). This *is* a Nash equilibrium in strategies. If Newvel chooses to Enter and if Microhard has adopted the strategy, Accommodate, the associated outcome is a best response for both Newvel and Microhard. That is, if Microhard has adopted a strategy to Accommodate, then Enter is the best response for Newvel and if Newvel enters accommodating is a best response for Microhard. Therefore the combination (Enter, Accommodate) is a Nash equilibrium.

What about the combination (Stay Out, Fight)? It also satisfies the Nash definition. If Newvel chooses Stay Out, then the Fight strategy is a best response for Microhard, while if Microhard has chosen its Fight strategy, then Stay Out is a best response for Newvel. Therefore, (Stay Out, Fight) is also a Nash equilibrium in strategies. We leave it for the reader to show that the strategy combination (Stay Out, Accommodate) is not a Nash equilibrium.

Again, it is important to understand that a Nash equilibrium is defined in terms of strategies that are best responses to each other. In the second Nash equilibrium (Stay Out, Fight), Microhard never actually takes or implements a fighting action. Instead, it relies

Table 11.1 Payoff matrix for the Newvel-Microhard entry game

		Microhard	
		Fight	Accommodate
Newvel	Enter	(0, 0)	(2, 2)
	Stay Out	(1, 5)	(1, 5)

fully on the *threat* to do so as a device to deter Newvel from entering. The Nash equilibrium concept is not based on what actions are actually observed in the market place, but rather upon what thinking or strategizing underlies what we observe. This is what is meant when we say we need to define a Nash equilibrium in terms of firms' strategies.

It appears then that there are two Nash equilibria to this game. There is, however, something troubling about one of these, namely, the Nash equilibrium (Stay Out, Fight). It is true that if Microhard has fully committed itself to the strategy, "Fight," then Newvel's best strategy is "Stay Out." But Newvel might question whether such a commitment is really possible. By adopting the Fight strategy, Microhard essentially says to Newvel, "I am going to price high as long as you stay out, but, if you enter my market, I will cut my price and smash you." The problem is that this threat suffers a serious *credibility* problem. We already know that once Newvel has entered the market, taking action to fight back is not in Microhard's best interest. It does much better by accommodating such an entry. Consequently, Microhard does not have an incentive to carry out its threat. So, why should Newvel believe that threat in the first place?

What we have really just discovered is that any Nash equilibrium strategy combination based on non-credible threats is not very satisfactory. This means that we need to strengthen our definition of Nash equilibrium to rule out such strategy combinations. This is where the notion of subgame perfection, or a subgame perfect Nash equilibrium, becomes important. A Nash equilibrium is considered to be subgame perfect only if at that point in the game when a player is called upon to make good on a promise or a threat, doing exactly that and fulfilling the promise or threat is what would be the player's best response. In other words, if any promises or threats are made in one period, carrying them out is still part of a Nash equilibrium in a later period should the occasion arise to do so. If Microhard adopts a strategy that includes the threat of a fight if entry occurs, then if the strategy is subgame perfect it must be optimal for Microhard to fight in the event that Newvel enters, i.e., in the subgame set in period 2 after Newvel has made its entry decision. Because it is not, any outcome that includes Fight cannot be part of a subgame perfect Nash equilibrium.

Generally speaking, it is more difficult to test for subgame perfection using the matrix representation of the game. This is the reason we originally found two Nash equilibria in the game above. For dynamic games, we prefer instead to use an extensive or tree representation of the game.

The extensive form of a game is comprised of branches, nodes, and vectors of payoffs. The nodes are the decision points of the game. The extensive form of the entry game between Newvel and Microhard is shown in Figure 11.2. Here, the nodes are labeled N (Newvel) or M (Microhard), depending on which firm makes the move at that position. The

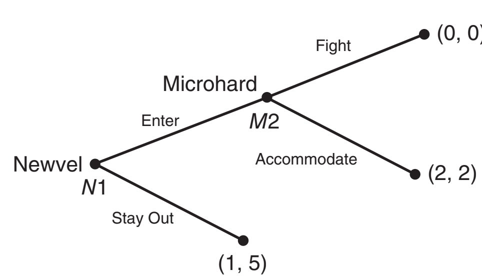

Figure 11.2 The extensive form of the Microhard-Newvel game

branches that are drawn from a node represent the choice of actions available to the player at that node. Each branch points either to another node, where further action takes place, or to a vector of payoffs (Newvel's payoff shown first), which means that this particular action has ended the game. Finally, at any node players know about the course of play that has led to that node.

When we represent a sequential game in extensive form it is easy to identify a subgame. A subgame is defined as a single node and all the actions that flow from that node. In the extensive game illustrated in Figure 11.2, there are two subgames. There is the full game starting from node $N1$ (the full game is always a subgame). Then there is the subgame starting at node $M2$, and including all subsequent actions that flow from this node. A strategy combination is subgame perfect if the strategy for each player is a best response against the strategies of the other players for every subgame of the entire game. In the case at hand, it is readily apparent that for the subgame beginning at node $M2$, the best response strategy for Microhard is Accommodate and *not* Fight. Hence, the strategy combination (Stay Out, Fight) cannot correspond to a subgame perfect equilibrium. The only such equilibrium in this case is that of (Enter, Accommodate).

There is an important technique for solving games with a finite number of nodes. In such games, the simplest way to identify the subgame perfect equilibria is to work backwards, eliminating branches that will not be taken until we have reduced the game tree to having a single branch from each node. This takes advantage of the property that a subgame perfect equilibrium strategy combination must be a Nash equilibrium in each subgame. In our example, we start at node $M2$. We have already seen that we can eliminate the "Fight" branch, leaving only the single "Accommodate" branch from node $M2$. Now pass down the tree to node $N1$. Newvel now knows that Stay Out leads to a payoff of 1, while Enter leads to $M2$ and to Accommodate by Microhard, giving Newvel a payoff of 2. So the Stay Out branch can be eliminated. The game tree now has a single branch from $N1$ and a single branch from $M2$, so we have solved the game. Newvel chooses to Enter and Microhard chooses to Accommodate. In other words, this procedure has eliminated the combination (Stay Out, Fight) as a perfect Nash equilibrium.

11.3

Practice Problem

Centipede is a well-known variant of games involving a chance to "grab a dollar." The game is played as between two players, as follows. A neutral third party, call it Nature, puts $1 on the table. Player 1 can either "grab" this dollar or "wait." If player 1 takes the dollar, the game is over and player 1 gets $1 and player 2 obviously gets nothing. However, it is completely understood that, if player 1 waits, Nature will *triple* the amount on the table to $3. At that point, it becomes player 2's turn to move. Player 2's options are as follows: either take the entire $3 or share the money equally with player 1.

a. Construct the 2×2 payoff matrix for this game taking player 1's actions to be either Grab or Wait, and player 2's actions to be either Grab (the whole $3) or Share. Assume the payoffs are equal to the amount of money the player receives.
b. Draw the game in its extensive form.
c. Suppose that player 2 promises player 1 to choose Share if player 1 chooses Wait. Is this promise credible? Why or why not?

11.4 THE CHAIN STORE PARADOX

In the Microhard and Newvel game, there is just one market and one potential entrant, and fighting the entrant was not an optimal response to entry. However, what if Microhard faced more than one entrant? Perhaps fighting one entrant builds a reputation for aggressive behavior that will scare off later entrants. The consideration of the reputational effects of fighting may change Microhard's optimal strategy. Taking predatory action against a rival—costly though it is—could be useful *if* it serves to make the threat credible against *other* rivals, either those in other markets or those who may appear later in time. If we introduce this possibility into our example, could Microhard's threat to fight become credible because the subsequent gains in other markets from establishing a reputation as a fighter are sufficiently large? In other words, could reputation effects make Fight credible and the strategy combination (Stay Out, Fight) subgame perfect?

The fact that extension of the above game to many markets (distributed over time or space) and to other rivals may *not* lead to a different outcome is a famous result dubbed by Selten as "The Chain-Store Paradox."[5] To see the logic of this puzzling result, consider a situation in which Microhard has established operating units in each of twenty markets, perhaps twenty different cities. In each city, Microhard faces potential entry by a single, small competitor. At the moment, none of these potential competitors has the capital to start operations. However, as time goes on, one after another will raise the necessary funds. To make matters simple, assume that the pay-offs in each of the twenty markets are just as in the pay-off matrix of the previous section. The question facing Microhard is how to react to this sequence of potential entrants. In particular, should Microhard adopt an aggressive response to the first entrant and drive it out of business? Will this tactic earn Microhard a reputation for ruthlessness such that subsequent entrants in its other markets will get the message and choose not to enter?

Again, working backwards can help us identify a subgame perfect strategy. So, let's start with one possible scenario in which Microhard is facing the last potential entrant in the final, twentieth market. It is possible that Microhard has followed through on its threat to cut price and drive out any entrant not just in the first market, but also in all previous nineteen markets. This is a possible path in the game and we are interested if such an aggressive response to entry can convince the last potential entrant to stay out, so that Microhard would be spared a fight in this final case.

However, consider the viewpoint of the entrant to the twentieth market. This firm will realize that because there are no subsequent entrants, it is playing a game that is exactly the one-period game we discussed above. Hence, this last entrant will understand that Microhard has no incentive to Fight in this last market. Microhard's profit is greater if it follows a "live and let live" strategy in this last case[6] because it cannot gain from any further demonstration of its ruthlessness. There are no other entrants left to impress! In other words, (Enter, Accommodate) must be the Nash equilibrium for this final subgame.

One might think that just because Microhard cannot credibly deter entry in the twentieth market, it is still possible to deter the entry of earlier rivals by means of the threat to Fight. To see why this is not possible, however, consider the potential entrant in the nineteenth

[5] Selten (1978). We have obviously limited ourselves here to consideration of finitely repeated games only. Infinitely repeated games are considered in the next chapter.

[6] Implicit here is the presumption that accommodating an entrant is in the short run more profitable than engaging in a price war.

rather than the twentieth market. Once again, let's take the extreme case in which Microhard has taken predatory or fighting action in the prior eighteen markets. Now the potential entrant in the nineteenth market can reason as well as we can. As a result, this firm will work out the logic of the preceding case and rightfully conclude that Microhard will not fight in the twentieth market. The entrant in the nineteenth or next-to-last market will then reason as follows: "Microhard will let the last rival firm survive because it is more profitable to do so. Yet this implies that both Microhard and I know that the entry of the last rival will not be challenged no matter what happens in this, the nineteenth market. It follows then that there is no reason for Microhard to act tough here. Its only reason to do so would be to convince the entrant in the next market. Because this is not possible, the only justification for fighting in market nineteen has been removed." Once again, Microhard's promise to fight is not credible. It gains Microhard nothing by way of a demonstration to the next rival. Absent such a reputation effect, Microhard's best response to entry in the nineteenth market is again to accommodate. Knowing this, the potential entrant in market nineteen will enter.

We can continue in this fashion repeatedly, bringing us back all the way to the initial market. At every stage, we will find that a strategy to fight after entry occurs is not in Microhard's interest. Accordingly, any strategy that includes the threat to Fight if entry occurs is not credible and therefore not subgame perfect. This is true at each node, from that at the twentieth market to that at the nineteenth market, and so on all the way back to the complete game starting with the very first market. Within this simple model, there is no way for the incumbent to credibly threaten an aggressive low-price response to entry.

If this were the end of the story, our interest in the predatory conduct would certainly be very low. Why should we worry about an event that presumably never occurs? The answer is that there may be ways to make the threat to fight credible other than actual fighting, itself.[7] In the next two chapters we will examine tactics by which firms may make their predatory threats credible and deter entry significantly.

11.5 EMPIRICAL APPLICATION: STACKELBERG BEATS COURNOT

The oligopoly models studied in this and the previous chapters—Cournot, Bertrand, and Stackelberg—are just that: They are models and therefore simplified abstractions of a very complicated reality. Often students learning these models ask whether anyone really thinks this way, i.e., whether anyone really works through the models to derive their optimal choices in the way game theory implies.

It is difficult of course to understand or to prove exactly what happens inside someone's mind. However, we can make intelligent inferences about their thought process by observing their behavior in different settings. A relatively recent method for doing this is to conduct experiments with real people in a controlled setting. An example of such work is the paper "Stackelberg Beats Cournot: On Collusion and Efficiency in Experimental Markets" by Huck, Müller, and Normann (2001).

In brief, Huck, Müller, and Normann (2001) divide 134 students into pairs to play simulated Cournot and Stackelberg two-person games. In each case, each of the two student players is told that they are one of two firms, each of which must choose how much to

7 Schelling (1960) contains early and lasting contributions to developing equilibrium notions for dynamic games. See also Tirole (1988) and Rasmusen (2007).

produce of an identical product at a constant marginal cost of $c = 6$, and facing the inverse market demand curve:

$$P(Q) = P(q_1 + q_2) = 30 - Q; \text{ for } Q \leq 30; \text{otherwise } 0 \tag{11.17}$$

In the Stackelberg games, one student is assigned the role of follower and plays only after observing the student assigned the leader role. In the Cournot games, each student chooses without knowing the choice of the other so that play is effectively simultaneous. Before the start of any game, players were also given a payoff table showing for each and every possible strategy combination the profit to each firm as measured in a fictitious currency called a "Taler." A single game between two players actually consisted of ten rounds in which the players would make the output choices ten successive times to allow for learning. Player motivation to maximize profit was provided by the fact that players knew that two of the ten rounds would be chosen at random and that they would then receive real money (Euros) at the rate of one unit for every ten Talers of profit. In the results below, we focus on the outcomes for games in which the rival players were assigned randomly before the start of each complete game.

Given the demand and cost structure, you should recognize that the standard Stackelberg outcome is for the leader to produce twelve units while the follower produces six, and that the standard Cournot outcome is for each firm to produce eight units. In turn, these outcomes also carry the implication that total output in the Stackelberg market will exceed that in the Cournot market. What happened in the laboratory experiments?

Table 11.2 shows the average outcomes across the sample of games along with the theoretical predictions for each scenario. A few findings are readily observable. First, the outcomes in the Cournot experiments are very close to those predicted by economic theory, with each Cournot student-firm typically choosing an output very close to eight units. Second, the Stackelberg experiments yield results that, while still preserving the greater output and first-mover advantage for the leader firm, depart somewhat from the Stackelberg theoretical model in that the leader's output is smaller than predicted while the follower's output is larger than predicted. Third, despite this latter finding, the typical Stackelberg game nevertheless still results in more total output than the typical Cournot game, again as theory predicts.

Huck Müller, and Normann (2001) explore the difference between the actual and theoretical Stackelberg predictions somewhat further by first comparing the best-response function implied by the actual players' behavior with that predicted by theory. The latter

Table 11.2 Actual and predicted outcomes for Stackelberg and Cournot experiments

	Stackelberg		*Cournot*	
	Actual Leader/Follower	*Predicted Leader/Follower*	*Actual Each Duopolist*	*Predicted Each Duopolist*
Individual Firm Output	10.19/8.32	12/6	8.07	8
Total Market Output	18.51	18	16.14	16
Total Combined Profit	93.48	108	116.60	128
Total Consumer Surplus	175.37	162	135.38	108
Total Welfare	268.85	270	251.98	236

is of course: $q_F = 12 - 0.5q_L$. Using linear regression, the authors find that the followers in their random pair experiments instead are best described as having a response function given by:

$$q_F = 10.275 - 0.178q_L \tag{11.18}$$

The slope of this estimated behavior is much flatter than predicted, implying that follower firms do not reduce their output as much as predicted in response to an output choice by the leader. For example, at the predicted leader output of 12, the response function in (11.18) implies a follower output of 8.35—well above the 6 units predicted by the theoretical best response function.

In short, Huck, Müller, and Normann (2001) find that followers are more aggressive than theory predicts. They hypothesize then that this is because followers have an *inequality aversion* in that they care not just about their own profit but also about the inequality between the profits of the leader and the follower. In particular, the authors suggest that instead of maximizing profit, follower players may instead be maximizing the following utility function:

$$U(\pi_F, \pi_L) = \pi_F - \alpha\ \max(\pi_L - \pi_F, 0) - \beta\ \max(\pi_F - \pi_L, 0) \tag{11.19}$$

Here α and β are typically taken to be positive fractions. Thus, equation (11.19) implies that if the leader and the follower have the same profit, $\pi_F = \pi_L$, the follower player cares only about maximizing his or her own profit as in the conventional model. However, if either $\pi_L > \pi_F$ or $\pi_F > \pi_L$ the follower loses some utility. Hence, the follower will be willing to sacrifice some profit depending on how large α and β are. Usually, it is assumed that $\alpha > \beta$, implying that the follower cares more about the leader being the one to earn the greater profit.

Consider for example, the typical leader output of 10.19 found in the experiments. If this output is taken as given, then the follower firm faces an inverse demand curve of: $P = 19.81 - q_{\mathrm{a}}$ and therefore a marginal revenue curve of $\mathrm{MR} = 19.81 - 2q_F$. Equating this with its marginal cost of 6, the follower would maximize its profit at $q_F = 6.9$, implying that 12.91 is the price. However, this would mean that $\pi_L - \pi_F = 70.41 - 47.68 = 22.73$, and this inequality in profits would reduce the followers utility if $\alpha > 0$. Note though that the follower is in a position of doing something about this. If the follower increases output a bit, this will lower the price and reduce the leader's profit. It will also lower the follower's profit but not as much as the leader's precisely because the follower's output is increasing while the leader's output is, by assumption, given. Therefore, increasing its output will reduce the profit difference between the two firms.

Because β is not relevant when $\pi_L > \pi_F$, we can now write the follower's utility function as:

$$U(\pi_F, \pi_L) = (1 + \alpha)\pi_F - \alpha\pi_L \tag{11.20}$$

It follows that the change in utility ΔU satisfies:

$$\Delta U = (1 + \alpha)\Delta\pi_F - \alpha\Delta\pi_L \tag{11.21}$$

At the optimal point, the follower is just balancing the greater profit equality its output increase is bringing against the actual profit loss that increase causes. Hence, $\Delta U = 0$ at the optimum implying that α satisfies:

$$\frac{\alpha}{1+\alpha} = \frac{\Delta \pi_F}{\Delta \pi_L} \tag{11.22}$$

It is straightforward to show that at the observed experimental average outputs in the Stackelberg experiments, $\Delta \pi_F / \Delta \pi_L = 0.277$, i.e., an increase in the follower's output would lower its profit by about 28 percent as much as it would lower the leader's output. From this, we can infer that for this "representative" case, $\alpha = 0.3845$. In other words, the follower players felt that having each firm have the same profit was about 40 percent as important as having its own individual profit maximized. It is this rather sizable preference for equal profit outcomes that Huck, Müller, and Normann (2001) deem responsible for the more aggressive follower behavior they observe. Whether that preference also characterizes real-world business people as opposed to experimental student players is an open question. Even with that possible qualification, however, it is still notable how closely the experimental outcomes follow the theoretical predictions for the relative Stackelberg and Cournot outcomes.

Summary

Sequential market games are different from simultaneous ones. Moreover, the effect of changing from simultaneous to sequential play differs depending on whether the strategic variable of choice is quantity or price. The basic sequential quantity game, typically referred to as the Stackelberg model, confers a large advantage to the firm that chooses production first. In the linear demand and cost case, the first mover in a Stackelberg game produces the monopoly output. The follower produces only half this much. Prices are lower than in the basic Cournot model, but the large market share of the first mover gives that firm an increase in profit over what it would earn in the simultaneous production game.

In contrast, a sequential price game with differentiated products yields higher profits for both firms than either would earn if prices were set simultaneously. Moreover, in this case, it is the firm that sets price last that does best if products are horizontally differentiated. Sequential price games are thus an example of games that have a second-mover as opposed to a first-mover advantage.

How close theoretical models correspond to economic reality is always a question. However, experimental evidence relying on classroom "laboratory" test runs with students playing Cournot and Stackelberg duopoly games suggests that these players do respond largely in the manner predicted by these models. This is true even though the actual monetary reward earned by these students averaged only about \$9. If such small motivations are enough to get student players to largely duplicate the theoretical predictions, one might argue that actual business leaders for whom the stakes of "getting it right" are much higher might come that much closer to the hypothesized ideal behavior.

Crucial to any sequential game is the issue of commitment. How do firms establish themselves as leaders or followers? How can a firm commit to its output or price strategy in a way that a rival finds credible? This issue is best explored by considering the game in its extended form and identifying strategy combinations that are subgame perfect, i.e., strategies that call for actions at later points in the game in which those actions continue to be optimal when the time comes to take them, given the history of play up to that date.

Threats and promises of later punishments and rewards are particularly important in games in which one firm is trying to prevent another from entering its market (or perhaps trying to induce it to leave). The question again is whether such threats and promises can be made credible. If they can, then incumbent firms may be able to maintain their dominant position in an industry and not fear competitive entry. This is the subject of our next chapter.

Problems

1. Consider a Stackelberg game of quantity competition between two firms. Firm 1 is the leader and firm 2 is the follower. Market demand is described by the inverse demand function $P = 1000 - 4Q$. Each firm has a constant unit cost of production equal to 20.
   a. Solve for Nash equilibrium outcome.
   b. Suppose firm 2's unit cost of production is $c < 20$. What value would c have so that in the Nash equilibrium the two firms, leader and follower, had the same market share?

2. Let's return to Tuftsville (Chapter 10) where everyone lives along Main Street, which is 10 miles long. There are 1,000 people uniformly spread up and down Main Street, and every day they each buy fruit smoothie from one of the two stores located at either end of the street. Customers ride their motor scooters to and from the store, using \$0.50 worth of gas per mile. Customers buy their smoothies from the store offering the lowest price, which is the store's price plus the customer's travel expenses getting to and from the store. Ben owns the store at the west end of Main Street and Will owns the store at the east end of Main Street. The marginal cost of a smoothie is constant and equal to \$1 for both Ben and Will. In addition, each of them pays Tuftsville \$250 per day for the right to sell smoothies.
   a. Ben sets his price p_1 first and then Will sets his price p_2. After the prices are posted consumers get on their scooters and buy from the store with the lowest price including travel expenses. What prices will Ben and Will set?
   b. How many customers does each store serve and what are their profits?

3. In Centipede[8] there are two players: player 1 moves first, player 2 moves second. After at most two moves, the game ends. The game begins with \$1 sitting on a table. Player 1 can either take the \$1 or wait. If player 1 takes the \$1 the game is over, and player 1 gets to keep the \$1. If player 1 waits, the \$1 quadruples to \$4. Now it is player 2's turn. Player 2 can either take the entire \$4 or split the \$4 evenly with player 1.
   a. Draw the extensive form for the game of Centipede.
   b. What is the equilibrium to this game? Can player 2's strategy of splitting the money ever be a part of an equilibrium outcome to the game?
   c. Now suppose that Centipede has three moves. Player 2 can now either wait, split the money, or take the \$4. If player 2 waits then the money on the table quadruples again and player 1 can either take it all or split it. Draw the extensive form for the new game and solve for the equilibrium outcome.

4. Dry Gulch has two water suppliers. One is Northern Springs, whose water is crystal clear but not carbonated. The other is Southern Pelligrino, whose water is naturally carbonated but also somewhat "hard." The marketing department of each firm has worked out the following profit matrix depending on the price per 2-gallon container charged by each firm. Southern Pellegrino's profits are shown as the first entry in each pair.

Southern Pelligrino's Price: \ *Northern Springs Price:*	3	4	5	6
3	24,24	30,25	36,20	42,12
4	25,30	32,32	41,30	48,24
5	20,36	30,41	40,40	50,36
6	12,42	24,48	36,50	48,48

   a. What is the Nash equilibrium if the two firms set prices simultaneously?
   b. What is the Nash equilibrium if Northern Springs must set its price first and stick with it, and Southern Pelligrino is free to respond as best it can to Northern Springs's price?
   c. Show that choosing *price* first is a disadvantage for Northern Springs. Why is this the case?

[8] This game was first introduced by Rosenthal (1981).

5. Suppose that firm 1 can choose to produce good A, good B, both goods, or nothing. Firm 2, on the other hand, can produce only good C or nothing. Firms' profits corresponding to each possible scenario of goods for sale are described in the following table:

Product Selection	*Firm 1's Profit*	*Firm 2's Profit*
A	20	0
A,B	18	0
A,B,C	2	−2
B,C	−3	−3
C	0	10
A,C	8	8
B	11	0

   a. Set up the normal form game for when the two firms simultaneously choose their product sets. What is the Nash equilibrium (or equilibria)?
   b. Now suppose that firm 1 can commit to its product choice before firm 2. Draw the extensive form of this game and identify its subgame perfect Nash equilibrium. Compare your answer to (a) and explain.
   c. The game is like the one in (b), only now suppose that firm 1 can reverse its decision after observing firm 2's choice and this possibility is common knowledge. Does this affect the game? If so, explain the new outcome. If not, explain why not.

6. Find three examples of different ways individual firms or industries can make the strategy "This offer is good for a limited time only" a credible strategy.

7. The Gizmo Company has a monopoly on the production of gizmos. Market demand is described as follows: at a price of $1,000 per gizmo, 25,000 units will be sold whereas at a price of $600, 30,000 will be sold. The only costs of production are the initial sunk costs of building a plant. Gizmo Co. has already invested in capacity to produce up to 25,000 units.
   a. Suppose an entrant to this industry could capture 50 percent of the market if it invested in $10 million to construct a plant. Would the firm enter? Why or why not?
   b. Suppose Gizmo could invest $5 million to expand its capacity to produce 40,000 gizmos. Would this strategy be a profitable way to deter entry?

References

Gabszewicz, J., Pepall, L., and J.F. Thisse. 1992. "Sequential Entry with Brand Loyalty Caused by Consumer Learning-By-Doing," *Journal of Industrial Economics* 60 (December): 397–416.

Goolsbee, A., and A. Petrin. 2004. "The Consumer Gains from Direct Broadcast Satellite and Competition with Cable TV," *Econometrica* 72 (March): 351–81.

Huck, S., W. Müller, and H. T. Normann 2001. "Stackelberg Beats Cournot: On Collusion and Efficiency in Experimental Markets," *The Economic Journal* 111 (October): 749–65.

Lieberman, Marvin. B., and David. B. Montgomery, 1988. "First Mover Advantages," *Strategic Management Journal* 9 (Summer): 41–49.

Rasmusen, E. 2007. *Games and Information*, 4th ed. Cambridge, MA: Basil Blackwell, Inc.

Rosenthal, R. W. 1981. "Games of Perfect Information, Predatory Pricing, and the Chain Store Paradox," *Journal of Economic Theory* 25 (August): 92–100.

Schelling, T. 1960. *The Strategy of Conflict*, Cambridge Mass.: Harvard University Press.

Selten, R. 1978. "The Chain Store Paradox," *Theory and Decision* 9 (April): 127–59.

Stackelberg, H. von, 1934. *Marktform und Gleichgewicht*, Berlin and Vienna. English translation by A.T. Peacock, (1952) London: William Hodge.

Sutton, J. 1991. *Sunk Cost and Market Structure*, Cambridge, MA: The MIT Press.

Tirole, J. 1988. *The Theory of Industrial Organization*, Cambridge: MIT Press.

Part Four
Anticompetitive Behavior and Antitrust Policy

The chapters in this section build on the game theoretic analysis of the previous three chapters to explore the tactics that firms can employ to blunt competitive pressures and thereby earn supracompetitive profits. In Chapter 12, we consider various tactics that exploit the first-mover advantage of the incumbent relative to a new entrant, such as the excess capacity investment stressed by Dixit (1980) and the bundling techniques discussed earlier in Chapter 8. We also present some empirical evidence regarding the use of excess capacity based on the Conlin and Kadiyali (2006) study of the Texas hotel industry.

In Chapter 13, we extend the analysis of so-called predatory behavior to include tactics that exploit either the information advantage that the incumbent has, as in the classic Milgrom and Roberts (1982) limit pricing paper, or the disadvantage that a small entrant might have vis-à-vis its creditors, as in the paper by Bolton and Scharfstein (1990). We then examine the use of long-term contracts to "tie up" consumers and deny a new entrant the customer base necessary to exploit scale economies. The empirical application for this chapter is based on the Ellison and Ellison (2011) study of the advertising decisions of pharmaceutical firms faced with potential new generic competitors as their key patents expire.

Finally, in Chapter 14, we consider the ability of firms to collude and suppress competitive pressures. Such price-fixing is a major concern of the antitrust laws. Indeed, the last fifteen years have witnessed the successful prosecution of a record number of price-fixing cartels. We begin with an analysis of the difficulties that must be overcome by cartel members if collusion is to succeed. We then show how collusion can become an equilibrium outcome in games of indefinite duration. This gives rise to the well-known Folk Theorem describing the conditions under which those obstacles can be overcome. This permits a discussion of antitrust policy including the recent practice of granting leniency to the first cartel members who cooperate with the authorities, and how important this has been in breaking international cartels. The chapter concludes with a discussion of Kwoka's (1997) empirical analysis of the impact of a real estate cartel that rigged auctions in the Washington, D.C., area.

12

Entry Deterrence and Predation

Elementary textbook presentations of microeconomics, such as our presentation in Chapter 2, often represent monopoly power as a transient phenomenon. The argument is that whenever a firm acquires market power and earns super-normal profit, entry will occur and the new entrants will reestablish a competitive market. Those who follow actual markets, however, know that this scenario may be more the exception than the rule. Campbell's, for example, has held a dominant position with 60 percent or more of the US canned canned soup sales for over a hundred years. For twenty years, Microsoft's *Windows* has maintained a share of over 90 percent in the market for personal computer operating systems, although this share drops to closer to 80 percent if tablet computers are included.[1] Despite recent expansion by Chinese and other emerging economy firms, Sotheby's and Christie's have together controlled roughly 55 percent of the world's fine art auction market for two decades, and as much as 70 percent of that market during many of those years.[2] Such instances of sustained dominance are common. Moreover, such anecdotal evidence is buttressed by formal analysis such as that by Baldwin (1995) and Geroski and Toker (1996). These authors find that, on average, the number one firm in an industry retains that rank for somewhere between seventeen and twenty-eight years. In short, there is abundant evidence that, in contrast to the textbook analysis, market power is lasting. This raises the question then as to how such firms can sustain their profit-winning position. Why don't new rivals emerge to compete away that market share and profit? Are there strategies that the dominant firms can adopt to prevent this from happening? If so, what are these strategies and what are their implications for market outcomes?

The questions just raised are the focus of this chapter. They are of much more than mere academic interest. The potential for large incumbent firms to eliminate or prevent the entry of rivals was a major motivation behind the enactment of antitrust laws in the first place. That concern has persisted through many antitrust cases ever since, including the dramatic Microsoft antitrust case in 2000.[3] Section 2 of the Sherman Act deems it illegal to "monopolize or attempt to monopolize . . . any part of the trade or commerce." It follows

[1] Gartner Press Release, "Windows 7 Will Become Leading Operating System in 2011," http:/www.gartner.com/it/page.jsp? id=17626149 (August, 2011).

[2] Art Market Monitor, "Art World Expansion Erodes Sotheby's/Christie's Market Share," http://artmarketmonitor.com (June 23, 2011).

[3] *United States v. Microsoft Corp*., 87 F. Supp. 2d 30 (D.D.C. 2000).

that enforcement of this provision requires an understanding of what a firm can do in order to "monopolize" the market.

Employing tactics that are only profitable if they deter rival firms from competing in a market is what economists call *predatory* conduct.[4] A firm engaging in predatory conduct wants to influence the behavior of its rivals—either those currently in the market or those thinking of entering it. Predatory conduct often involves the making of threats and, if necessary, actually implementing the threats in a way that ensures such threats are *credible*. Credibility is absolutely essential for predatory conduct to be successful. After all, as we learned from the Chain Store Paradox in the last chapter, "talk is cheap." A threat aimed at dissuading a rival from entering one's market will only have the desired effect if it is credible. Such threats work when the rival or prey believes that the predator really "means business" and will pursue the predatory conduct *if the rival chooses to ignore the threat*.

In this chapter, we examine conduct designed either to deter rivals from entering an incumbent's market or to induce existing rivals to exit. Fear over such conduct was a fundamental motivation behind the initial antitrust provisions, which focused on efforts "to monopolize . . . any part of the trade or commerce" and to "materially reduce competition."[5] The basis for this legislative concern is, of course, the fear that with existing rivals and the threat of entry removed, a dominant firm will pursue monopoly practices that reduce efficiency. In this chapter, we limit ourselves to cases of complete information. We defer the examination of predatory conduct under incomplete information or uncertainty until the next chapter.

12.1 MARKET STRUCTURE OVER TIME: RANDOM PROCESSES & STYLIZED FACTS

Before we begin to model incumbent firm strategies to deter rivals' entry and encourage rivals' exit, we need to consider more fully what theory and evidence can tell us about how an industry's structure might evolve over time. As a simple example, consider an industry comprised of 256 firms, each with sales of $10 million. Suppose that in any period, sales at each firm will decline by 30 percent with probability 0.25, stay the same with probability 0.5, or return to their prior level with probability 0.25 for any firm of any size. (If these rates of growth or decline seem large, think of a period as say three or four years long.) Table 12.1 shows the evolution of firm size over just four periods. We can see that even over this short amount of time, the distribution of firm sizes is becoming very unequal. The largest firm is now nine times as large as the smallest and the top nine firms account for roughly the same amount of output as the bottom forty. This inequality will grow even larger as we let the process run over more periods of time.

The trend toward increasing inequality is a common feature of any process in which growth in any period is a random variable independent of firm size and previous growth rates. It is broadly referred to as Gibrat's Law after Robert Gibrat (1931) who made the following formal argument. Let X_t be a firm's size at time t. Let this be related to its size at $t - 1$ by the following stochastic process:

$$X_t = (1 + v_t)X_{t-1} \tag{12.1}$$

4 See, for example, Fisher (1991).

5 Our definition is also similar to that of Ordover and Willig (1981).

Table 12.1 Size distribution of firms after four years starting with 256 equal-sized firms

	Sales (Millions)								
	\$2.40	\$3.43	\$4.90	\$7.00	\$10.00	\$13.00	\$16.90	\$21.97	\$28.56
Period 0	0	0	0	0	256	0	0	0	0
Period 1	0	0	0	64	128	64	0	0	0
Period 2	0	0	16	64	96	64	16	0	0
Period 3	0	4	24	60	80	60	24	4	0
Period 4	1	8	28	56	70	56	28	8	1

Here v_t is a random disturbance term that we assume is normally distributed and has mean μ and variance σ^2. Next, take the natural log of both sides. If the time interval between t and $t-1$ is short, we may use the approximation that $\ln(1+v_t) \approx v_t$, where we uses lower case letters to denote logs. Hence, we may now write:

$$x_t = x_{t-1} + v_t \tag{12.2}$$

Denoting x at time $t=0$ as x_0, repeated substitution yields:

$$x_t = v_t + v_{t-1} + v_{t-2} + v_{\mathrm{t}-3} + \ldots + x_0 \tag{12.3}$$

This last equation says that the logarithm of the firm's size at time t will just be a random variable reflecting the accumulation of all the random growth shocks it has experienced up to that time. Each shock is itself a random variable drawn from the normal distribution. Thus over k time periods, as the importance of x_0 approaches zero, x_t will converge to a random variable with mean $k\mu$ and variance $k\sigma^2$. In turn, this implies that the variance of firm size itself will be $e^{k(2\mu+\sigma^2)}(e^{k\sigma^2}-1)$, which of course is far larger than $k\sigma^2$. Indeed, over a long time period the sum of all those accumulated shocks is also a normal random variable. Recall however that logarithms reflect exponential power. As the log of a firm's assets doubles, the actual volume of those assets is squared. So, although the log of firm size may be normally distributed, the distribution of actual firm sizes will be skewed. Those firms with above average values for the log of firm size will have many times above average values when size is measured without logs.

As mentioned previously, the outcome described above in which a market concentration increases endogenously over time is usually referred to as Gibrat's Law after its originator Robert Gibrat (1931). It has generated enormous interest in part because it suggests that oligopoly or even monopoly are natural market outcomes. In a rough way, it is consistent with the evidence cited earlier that dominant firms retain their market power for a considerable length of time. In fact, in an early study motivated by Gibrat's work, Kalecki (1945) produced some evidence that was very supportive of this natural concentrating tendency.

To a large extent, however, the subsequent research on Gibrat's Law focuses primarily on what that analysis leaves out rather than what it keeps in. This is because the Gibrat process is very mechanistic. There is no talk of research and cost-saving innovations that enhance a firm's growth. There is no consideration of mergers and firm combinations over time. Perhaps most relevant for our present purpose, there is no discussion of new firms entering an industry or older firms leaving, or what strategic interaction may lie behind such entry and exit. Subsequent research has tried to remedy these omissions and to develop theoretical models of industry evolution that build in these features [See, for example, Jovanovic (1982), Nelson and Winter (1982), Sutton (1997), and Klepper (2002).]

Of course, any theoretical model must ultimately confront the real world facts. Over the last twenty-five years, economists have worked hard to review the data and to document any empirical regularities or stylized facts that appear. A survey of these studies suggests that there are four stylized facts that any theory of industrial evolution must explain.

The first is that *entry is common*. Dunne et al. (1988, 1989), using US census data between 1963 and 1982, computed rates of entry in a wide cross section of two-digit SIC manufacturing industries. Their estimate of the average entry rate in manufacturing—defined as the number of *new* firms over a five-year period relative to the number of incumbent firms at the start of that period—ranged between 41.4 percent and 51.8 percent (about 8 to 10 percent on an annual basis). For the United Kingdom, Geroski (1995) estimated somewhat smaller but still significant annual rates of entry for a sample of 87 three-digit manufacturing industries. These ranged between 2.5 percent and 14.5 percent over the period 1974 to 1979. Cable and Schwalbach (1991) show that similar rates of entry exist across a wide range of developed countries. More recently, Jarmin et al. (2004) showed that rates of entry are even higher in the retail sector and may reach well over 60 percent, especially during periods of economic prosperity.[6]

The second stylized fact is that when entry occurs it is, by and large, *small-scale entry*. The studies by Dunne et al. (1988, 1989) showed that the collective market share of entrants in an industry ranged between 13.9 and 18.8 percent, again over a five-year interval.[7] Similarly, in Geroski's (1995) UK study, the market share of a new entrant was found to be quite modest, ranging from 1.45 to 6.35 percent. In the United States, Cable and Schwalbach (1991) find that while new entrants typically constitute 7.7 percent of an industry's firms in any year, they account for only 3.2 percent of its output. The typical share of entrants in retailing is noticeably larger, say closer to 25 percent, according to Jarmin et al. (2004), but they also find that this value has been declining over recent years.

The third stylized fact is that the *new entrant survival rate is relatively low*. Dunne et al. (1988, 1989) find that roughly 61.5 percent of all entrant firms exited within five years of entry and 79.6 percent exited within ten years. The corresponding exit rates found in retailing by Jarmin et al. (2004) are a very similar, between 59 percent and 82 percent. Birch (1987) used Dun and Bradstreet data for all sectors in the United States including, but not limited to, manufacturing and found that about 50 percent of all new entrant firms fail within the first five years.

Our final stylized fact that appears to hold in every study is that while rates of entry and exit vary across industries, *industries with high entry rates also have high exit rates*. In other words, entry and exit rates are strongly correlated. To take just one clear example, Cable and Schwalbach (1991) found that corresponding to an entry rate of 7.7 percent accounting for 3.2 percent of industry output, the exit rate is 7.0 percent and similarly it accounts for 3.3 percent of industry output. This finding is a little surprising because it does not appear consistent with the hypothesis that entry occurs in response to above-normal profit or that

6 The Dunne et al. (1988, 1989) entry (and exit) estimates are generally higher than those obtained by other researchers owing to the fact that they explicitly recognize the multiproduct and multiplant nature of firms.

7 Dunne et al. (1988) did find that existing firms who enter a new market through diversification typically enter at a larger scale than new, or *de novo*, entrants do.

exit reflects a below-normal profit. If profit is high, and therefore entry attractive, there is no reason for firms to leave. Similarly, if profit is so low that firms are induced to leave the industry, there ought to be little incentive for new entrants to emerge.

Taken together, the stylized facts reported above can be read as suggesting a sort of revolving-door setting in which mostly small firms enter, eventually fail, and exit, only to be replaced by a new cohort of small-scale entrants. In this view, the major difference across industries would be the pace at which this entry-fail-exit cycle proceeds. One interpretation of this evidence is that it reflects repeated attempts and, just as often, repeated failures of small firms to penetrate the markets dominated by large incumbents. This may help explain the correlation between entry and exit rates. Incumbents in markets that seem the most tempting targets may be those firms that fight the hardest against new entrants.

More formal support for this revolving door interpretation is offered by Urban et al. (1984) on the benefits of incumbency. They studied 129 frequently purchased brands of consumer products in twelve US markets and found that market shares were a decreasing function of the order of entry of the brand. Earlier entrants enjoyed larger market shares, all else equal. Similar results have been found by Lambkin (1988), Mitchell (1991), and Brown and Lattin (1994).[8]

A more recent analysis that combines an underlying theoretical random process with actual market data is Sutton's (2007) examination of forty-five Japanese industries over a twenty-three year period. Given the initial distribution of market shares and the variance of shocks to those shares, Sutton (2007) can calculate the expected number of initial industry leaders that will have lost their position due to random shocks after say fifteen or twenty years. In both cases, he finds that the actual number of such leadership losses is significantly less than what one would expect, i.e., market leaders retain their position significantly more than one would anticipate given the normal randomness those shares exhibit. Thus, this finding also suggests that leading firms retain their dominance longer than one would normally expect.

To be sure, the persistence of market dominance may well reflect the superior cost efficiency or higher product quality of those firms that emerge as market leaders early on. However, it is frequently alleged (especially by failed rival entrants) that the ability of early entrants to sustain a dominant industry position also reflects deliberate behavior aimed either at driving new entrants out or preventing their entry in the first place. This is the primary issue addressed in this and the subsequent chapter.

12.2 PREDATORY CONDUCT AND LIMIT PRICING

As noted above, there is a difference between a rational competitive response to rival entry and truly predatory behavior. We know, for example, that both the static Cournot and Bertrand models predict that when a second firm enters a previously monopolized market, the initial incumbent's response will lead to a (possibly dramatic) fall in price. While this may make the entry less profitable, it does not reflect anything other than a normal competitive response to competitive pressure. For true predation to occur, economists

[8] As Caves (1998) notes, though, there is regression toward the mean in firm growth rates. That is, really large firms tend to grow more slowly than do small ones. This feature blunts the ever-increasing concentration tendency implied by Gibrat's Law.