Chapter 4: Economic Principles Overview

"If it is the case that Amazon's model is better [than traditional retailing], it will get so much competition that margins will be forced down. If it is not so good, then it will not have competition, but it will not make much money."

Brett Trueman, Professor of Public Accounting, Haas School of Business, University of California at Berkeley, in *The Industry Standard*, June 30, 2000.

Prologue

In June 2000, a bond analyst at investment bank Lehman Brothers questioned the ability of Amazon.com, the online retailing world's 800 pound gorilla, to keep paying its bills without drastic improvement in its business. This brought a firestorm of criticism from Amazon boosters, including the company itself. Yet the most interesting question was sidetracked in the debate about the survivability of Amazon.com. Amazon's head start, its size, and its powerful brand are likely to pull it through, but it is not clear at all that it will ever generate anywhere near the kinds of profits that would justify the expectations of the more optimistic, as expressed in forecasts of an increasing stock price. Why is that? The quote from Brett Trueman, above, summarizes the skeptic's view, one that is firmly grounded in standard economic analysis.

Stepping back from the controversy entirely, we may ask what makes Amazon.com at all a likely challenger to established bricks-and-mortar firms such as bookseller Barnes & Noble, or discounting behemoth Wal-Mart? The obvious answer, of course, is that online shopping is convenient. It saves time and effort, and this gives Amazon an edge in selling to its 20 million customers. Yet if Amazon is to eventually make money, that convenience has to be worth enough to its customers so that Amazon can price at a level that covers its costs. How much is an individual's time worth? How can we set about getting a conceptual answer to that question, without expensive trial and error? Again, economic theory provides us with the tools we need to tackle this problem. Read on!

4.1 Introduction

An economist knows the price of everything and the value of nothing. So goes the old joke, with 'economist' substituted for the original 'cynic'. Yet one of the main threads in the development of economic ideas is understanding precisely the relationship between prices, as determined in the marketplace, and underlying subjective values of those who possess or consume the products and services bought and sold. The idea of value has also taken on a new life and meaning, with its use in the term 'value chain', describing the complex of activities that take place within the boundaries of a firm, and outside the formal marketplace.

Since the premise of this book is that economic principles can help us understand, and even predict, the tremendous changes occurring in modern economic life, we take a few pages to review, very briefly, the key economic principles that make up the core of what we now call microeconomics, but earlier economists would typically have called 'the theory of value' or 'price theory'.

We begin the review with firms, and their activities of producing and selling. We describe how the nature of the technology and the size of the market help determine the character of firms' activities in the market, in particular the degree of market power they have. We point out some of the things that are missing from the usual simple microeconomic story of what firms do, leaving these gaps to be filled in future chapters.

4.2 What Firms Do

The term 'firm', as used by economists, does not fit into a precise legal category. For an economist, a firm can be any entity that produces and sells a good or service. The exact boundary between a household and a firm can therefore be somewhat hard to draw. The seemingly increasing numbers of people running small businesses out of their homes, thanks to information and communication technology, is a modern illustration of this fuzzy boundary. Yet if we look back in history, the fuzzy boundary has always been there. Firms were originally not legal entities at all. Business meant family business, without the use of contracts among those involved. Production meant subsistence family farming, or artisanal activities, or cottage industry.

We can not review the fascinating history of the development of firms, for this is really the history of capitalism itself. We can encapsulate it, however, by noting that the history of firms is the history of the development of organizational forms, and organizational complexity. Even though, at one extreme, a firm may be a single person carrying out the full range of activities involved in making and selling a tangible product or providing an intangible service, the essence of the economic idea of a firm is the organization of this range of activities when carried out by more than one individual. To make this process work, objectives, obligations and rewards must be defined. For the economist, therefore, the firm is a 'nexus of contracts'.

Contracts may be implicit, as in the traditional family firm, where relationships and reputation rule. More often, the contracts that define a firm must include explicit legal contracts, among the parties involved, and between the members of the firm and the government. A firm can be formally constituted as a sole proprietorship, a partnership, or as a corporation (with many variants in the latter category, in particular). The parameters of each category define what the rights and obligations are of the owners of the firm, as well as the rights and obligations of the government (its right to tax, its obligation to enforce the contract among the firm's members). Formal legal contracts do not make relationships and reputation obsolete, but they are an essential part of the story of the rise of the modern firm: the legal structure of society enables and supports organizational complexity.

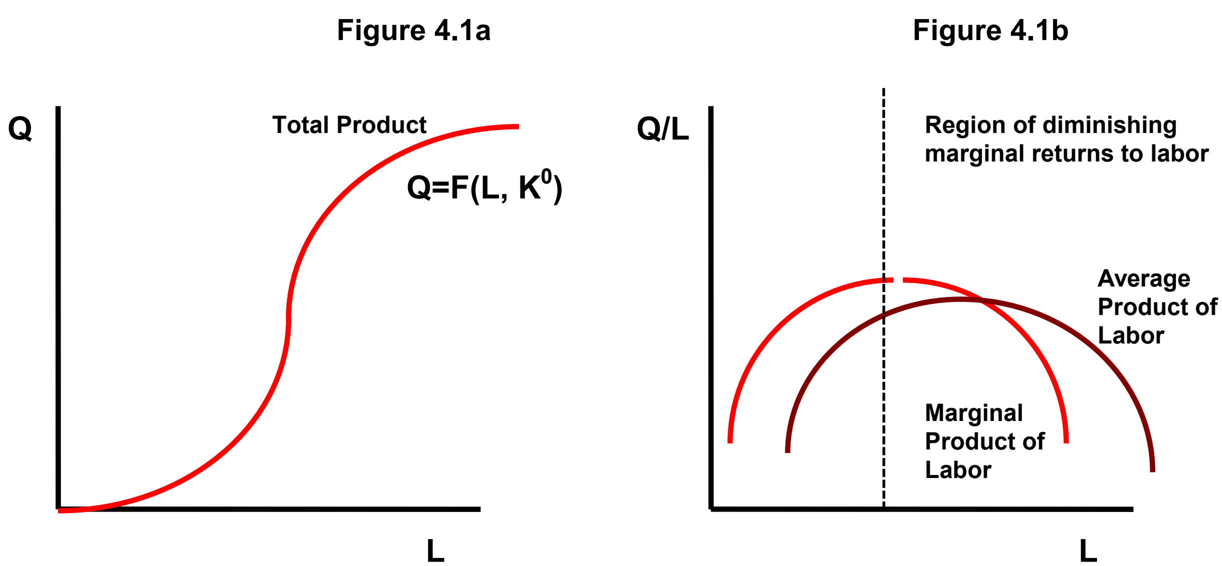
A huge chunk of economic analysis in the last three decades has concerned itself with the nature of firms, and their organization. Unfortunately, little of this work is reflected in intermediate microeconomics texts, or even in those that concern themselves with managerial economics (though business strategy texts have to face these issues squarely). One reason is the desire to keep things simple; another is to focus on the economic essence of production and selling. However, information technology broadly, and the Internet particularly, have profound implications for how firms are organized and how they operate. Therefore, the changes that are occurring and will occur are taken up throughout the book, but particularly in Chapters 5, 8 and 10. In doing so, we will also bridge the gap that typically exists between microeconomics and management views of the firm.

In this chapter, however, we provide a more stripped-down, basic look at firms, along the lines of standard undergraduate economic theory. Since we assume familiarity with the theory, the exposition here is quite brief. The essential ingredients of the theory are technology, input decisions based on the interaction of technology and input market conditions, and output and pricing decisions based on costs (the result of input decisions). The characterization of technology uses a black-box approach, where sequences of processes and material transformations are compressed into a single mathematical function, the production function. This is quite enough for a surprisingly powerful array of insights. Much of the action, from the point of view of economic understanding, comes from considering the nature of costs. The importance of the structure of a firm's costs, in particular how they relate to its size, is one of the key analytical insights for predicting what firms will do or should do, and what ultimately happens to them and the structure of the industry they are in.

Technology For present purposes, technology is described quite abstractly as a mapping of inputs to outputs. The inputs are typically characterized as labor and capital, but may be detailed as finely as we like in these categories, e.g., different kinds of machines and different skill levels and types of labor (software engineers, hardware engineers, marketing experts, accountants, and so on!). In addition, energy may also be included as an input category. This list excludes raw materials, so the result of the mapping, the output, when multiplied by price, is a value added concept. For simplicity, we assume the firm just produces one product or service.

Alternatively, we may include raw materials and intermediate inputs, and treat output as gross output. The value of this gross output is then what firms normally count as revenue. The arbitrariness of this revenue calculation as a measure of the economic activity of a firm should be noted, as revenue has become a focal point for e-commerce start-ups. Thus Priceline.com, the 'name-your-own-price' seller of airline tickets, hotel rooms, and so on, prefers to report include in its revenue the total value of the airline tickets it sells, since it argues that it possesses those tickets for a short time before the traveler receives them. Its own commission, which would be the appropriate measure of value added, is much lower. Economic realities and accounting conventions do not always align, as economists are always keen to point out.

The simple Q = F(L, K) production function as a stylized description of technology therefore fails to include the various steps in production that comprise the 'value chain' (Chapter 5) -- a key conceptual model for business strategists. It also neglects making the time dimension explicit. In fact, the production function represents a transformation over a fixed period of time, such as a week, month or quarter. Labor (L) and capital (K) in the formula are properly thought of as flows of services, and quantity of output (or, synonymously, total product, Q) is also measured in units per period. In this framework, an inability to speed up production would have to thought of as a case where an increase in the inputs is not possible because of market constraints, or as a situation where increasing the inputs does not result in an output increase (because of some other implicitly constrained input).



The second situation is an extreme case of diminishing returns. The idea of diminishing returns is simple and powerful, and is an important consequence of the simple production function framework, despite all its limitations. The so-called law of diminishing returns says that the marginal physical returns to any input will eventually decrease as the quantity of the input increases, if technology and other inputs are unchanged. This seems quite a reasonable statement, even if it is not a 'law' on par with the law of gravity. Diminishing returns are illustrated in Figure 4.1. The left-hand panel (4.1a) shows how total output might respond to increases in labor input, given a particular technology and capital input. The right hand panel (4.1b) shows average and marginal output. In particular, marginal product goes down after a point, as labor services are increased while capital services are fixed.

Illustration Box

Scale and Scalability

We can illustrate more formally the idea of constant returns to scale, and why considering knowledge more explicitly can imply increasing returns to scale. We can also relate these concepts to a term that is very common in discussing the information technology (IT) required for e-commerce, or e-business in general: scalability.

Consider once more the production function, the mathematical abstraction which summarizes how inputs are transformed into outputs. If the production function is denoted $Q = F(L,K)$, then constant returns to scale mean that doubling the amounts of labor and capital services will double output, so $F(2L, 2K) = 2F(L, K)$. This approach basically takes the technology as a given, captured somewhat mysteriously in the function 'F'. However, at least some of this technology, if not all of it, is embodied in knowledge that has tangible expression, and that can therefore be bought and sold. The most relevant example for us is software that governs production processes (recall the discussion of software patenting in Chapter 3). If we break software out as a separate input, then we can write the production function as $Q = G(L, K, S)$. We now have to use a different letter to denote some of the 'mysterious' aspects of production formerly hidden in the function F, because we are explicitly acknowledging the role of software. Now if labor and capital can be doubled, but the same software used (it doesn't have to be rewritten), then we should have $G(2L, 2K, S) = 2G(L, K, S)$. Since all inputs have not been doubled, but output has doubled, we really have increasing returns to scale: 'doubling' S as well, by writing better software, would presumably lead to output more than doubling. This fact was hidden when we failed to break out the knowledge input in production.

The situation described in the last equation is roughly what IT people mean when they talk about the 'scalability' of software. In practice, this scalability is not automatic. Capital includes hardware, and doubling the number of machines, all running the same software, may not handle double the amount of web traffic, for example. Additional software may be needed to allocate traffic among the machines, or perform other balancing acts. Thus increasing returns to scale in the economist's sense (implied by scalability in the IT sense) may hold only over a limited range. In the case of other kinds of knowledge, such as how to conduct a chemical reaction, this limitation may be less relevant, so increasing returns apply with full force.

While the traditional analysis emphasizes that returns eventually diminish, one of the ideas that seems to have permeated discussions of e-commerce is the presence of pervasive increasing returns. Certainly, if some input is indivisible (say, a blast furnace) then the returns to other inputs that are more divisible (such as labor) will increase as the amount of the inputs increases. This is the case in Figure 4.1, where there are initially increasing returns to labor. However, the notion of increasing returns that is alleged to matter for e-commerce is more complex than the simple technology story, which fits better the rise of large firms in the nineteenth century than it does the current situation. We will return to the idea of increasing returns due to 'network effects' in Chapter 16.

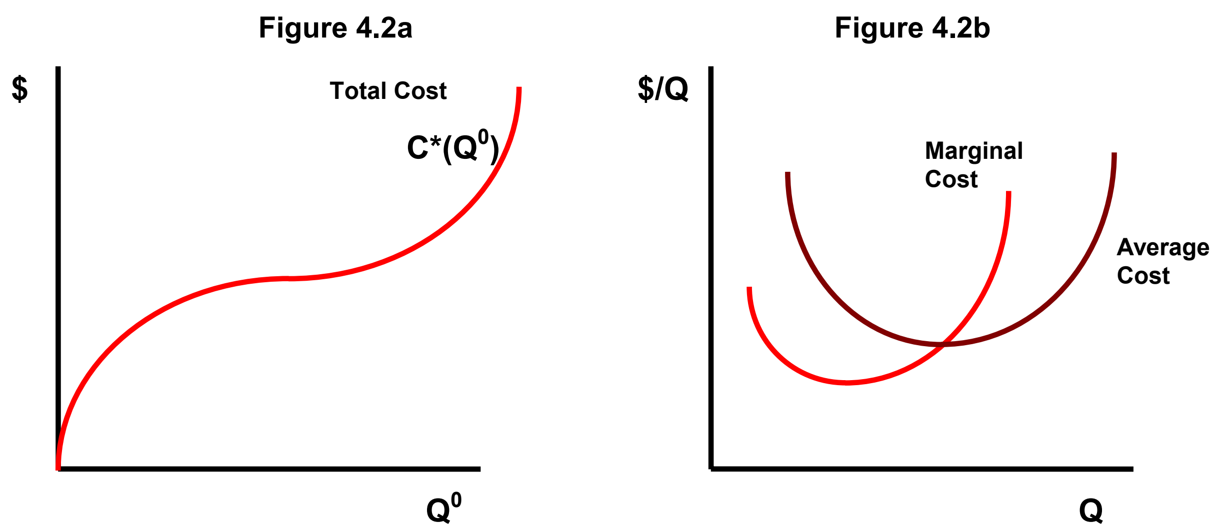
However, knowledge is one input that strengthens the argument for increasing returns that are closer to the classical kind. If knowledge is taken out of the production function, F, itself, and treated as a separate input (which also makes it easier to think systematically about innovation), then there is a case for pervasive increasing returns. Here we shift the concept of returns slightly, and consider the returns to altering the levels of all inputs together, not just one at a time. This is a scaling up of production, and we call this concept returns to scale. The presumption in this case is that replication of inputs, or proportionate increases, should lead to the same proportionate increase in output: this is what we call constant returns to scale. But this replication argument considers only physical inputs. The same input of knowledge can be applied over and over again. Hence, once knowledge is taken account of, there must be increasing returns to scale (see Illustration Box, above).

To summarize, almost the only feature of technology that we focus on at this level of economic analysis is the response of output to changes in input levels in any given period. What matters is whether the output response, as determined by technology, is less than proportionate, proportionate, or more than proportionate. To explain why this is what matters, we turn to considering costs.

Costs Input market conditions, together with technological possibilities, determine costs. While in practice a firm will choose its input levels based on how much it wants to produce to maximize profits, it is analytically convenient to separate the firm's input decisions and output decisions. We do this by looking at the firm's input decisions for any arbitrary output level.

Conceptually, this process can be described using the simple production function with labor and capital as inputs. We suppose that the firm is able to buy its inputs at going market prices, without its purchases affecting those prices. If w is the per period cost of labor and r is the per period cost of capital (including implicit or opportunity costs), then the firm's total cost is $wL + rK$. However, the firm is constrained to use enough labor and capital to produce at least an output level of Q^0, that is, $F(L, K) \geq Q^0$. The firm's choice variables are L and K, and the givens it faces are w, r and Q^0. Thus, if it makes its choices to minimize its costs, while respecting the inequality constraint imposed by the available technology, its optimal input choices will depend on the givens w, r and Q^0. If these choices are denoted $L^*(w, r, Q^0)$ and $K^*(w, r, Q^0)$, the firm's lowest cost of producing Q^0 is $wL^*(w, r, Q^0) + rK^*(w, r, Q^0)$. Note that this expression depends on w, r and Q^0. We can therefore denote it directly as a function of these quantities, say, $C^*(w, r, Q^0)$. This is what economists call the cost function.

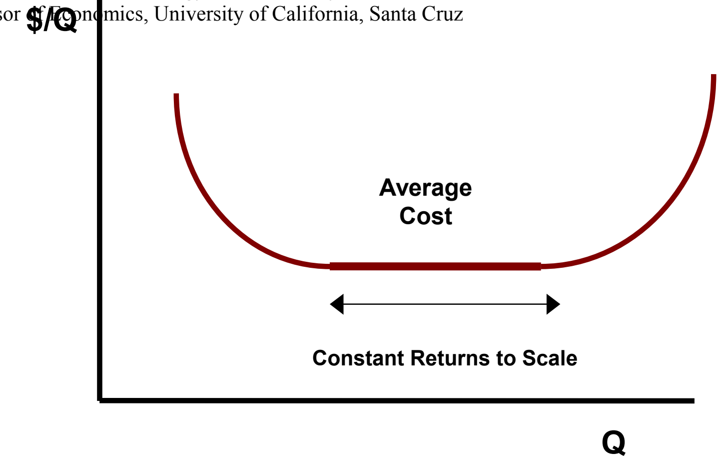
Note that we have not bothered to work out the characteristics of the cost-minimizing input choices. Those characteristics are not really central to those properties of the cost function that matter for our purposes. In fact, we are going to ignore the role of input prices for now, and write the cost function as $C^*(Q^0)$. The nature of returns to scale is closely tied to the behavior of cost as a function of output. If output increases more than proportionately as inputs increase, then costs will increase less than proportionately as the output obtained from those inputs increases. Thus we can just as well think of returns to scale in terms of the cost function. All the details of the technology can, for this purpose, be swept under the rug! The typical cost functions, familiar from introductory microeconomics texts, are shown below in Figure 4.2.



The left hand panel (4.2a) shows the cost function that we introduced earlier, now labeled the total cost, to distinguish it from the curves in the right hand panel (4.2b). The darker curve in that panel is the average cost, or simply the total cost divided by the quantity that is being produced. Simple geometry can be used to show that the shape of the total cost curve implies the U-shape of the average cost curve. (This is left for you to do as a review exercise.) The lighter curve is the marginal cost, which is the additional cost required to produce one more unit of output. Again, its shape is determined by the shape of the total cost curve, but a precise demonstration of that is somewhat harder (unless you use calculus). The fact that the marginal cost and average cost are equal precisely at the bottom of the U of the average cost curve is yet another fact that is mathematically guaranteed.

The shape of the average cost function is closely connected to the nature of the technology. In essence, the declining part of the average cost curve corresponds to a range of output where the technology is characterized by increasing returns to scale, and the increasing part of the average cost curve corresponds to decreasing returns to scale. Right at the bottom of the average cost curve is where there are constant returns to scale. If this bottom were flat, there would be a range of output over which constant returns will hold. This is illustrated in Figure 4.3.

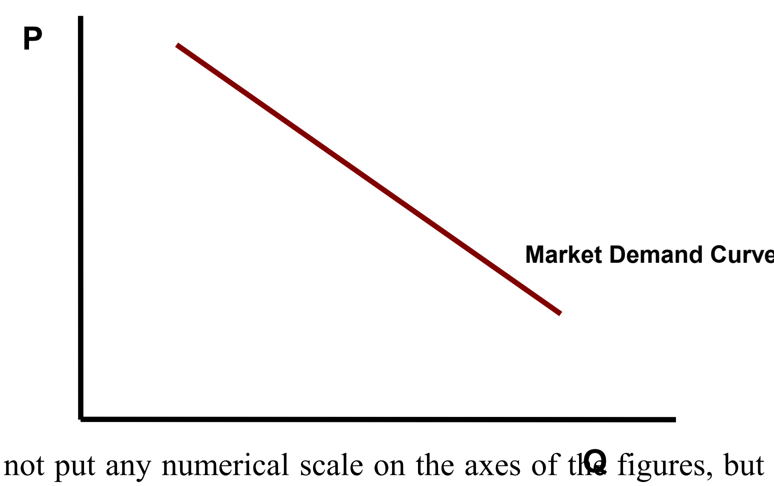
Figure 4.3



As we shall see next, how the shape of the cost curve stacks up against the size of the firm's demand is crucial for understanding how firms behave in the market, and how well they ultimately do. What is surprising is how much this simple model of firms can tell us, even about 'dot.coms' and other manifestations of the information age. The details about what firms do, that we have left out, are important, but we can sweep them aside for now and still get powerful predictions. That is the lesson of what follows.

Output and Prices Introductory economics courses emphasize the basic idea that the subject is the study of using scarce resources to satisfy human wants. But those wants have to backed up by the willingness and the ability to pay to have them satisfied. That willingness and ability to pay is captured geometrically in the demand curve for a product or service. We will review 'where demand curves come from' in Section 4.3, but for now assume that you are familiar with the idea of a downward-sloping demand curve, as shown in splendid isolation in Figure 4.4. Here 'P' stands for price, and 'Q' for quantity. Thus the two axes are measured in the same units as the axes for the average cost curve.

Figure 4.4



We have not put any numerical scale on the axes of the figures, but now relative scales will be important. Consider Figure 4.5, where the market demand curve is the same in the two panels, but the technology is very different with respect to the level of output at which decreasing returns to scale set in. In the left hand panel (4.5a), decreasing returns set in at a level of output that is low relative to the quantity demanded in the market at prices close to average cost. The firm is going to be 'small relative to the market'. Two consequences follow from this:

1. There is room for many other firms in the industry.
2. A single firm will have little or no market power: by itself, its output decisions will have little or no impact on the market price.

The market size and the technology, as shown in Figure 4.5a, together will tend to make the industry structure relatively competitive. This competitive structure dictates how output and prices will be determined in this industry/market.

The right hand panel (4.5b) shows the same market size, but with a very different technology. Now there are increasing returns to scale throughout the range of output that matters for this market. While there might be room for more firms in this market, it will be at the expense of higher average costs, and there will be a tendency for any of those firms to expand output and undercut their rivals. The industry/market in this case will tend towards being a monopoly (one seller) or an oligopoly (few sellers), and the firm or firms that operate in it have market power, in the sense that their output decisions will influence the market price, whatever other firms do. The existence of increasing returns to scale relative to the size of the demand tends to keep the number of firms in the industry small, and preserves their market power.

Figure 4.5a

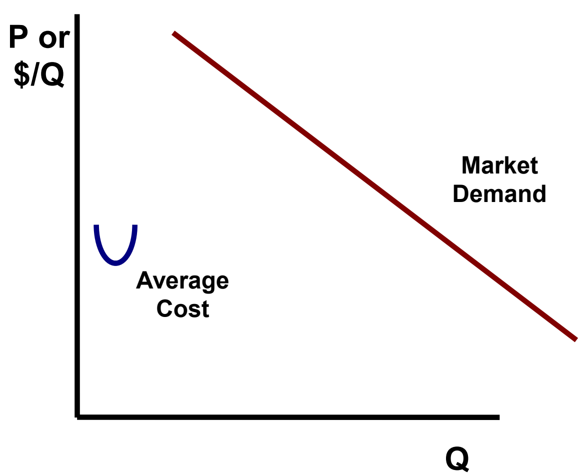
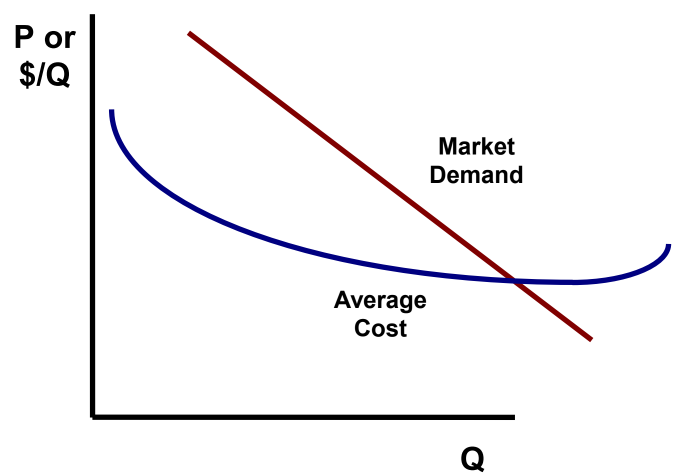
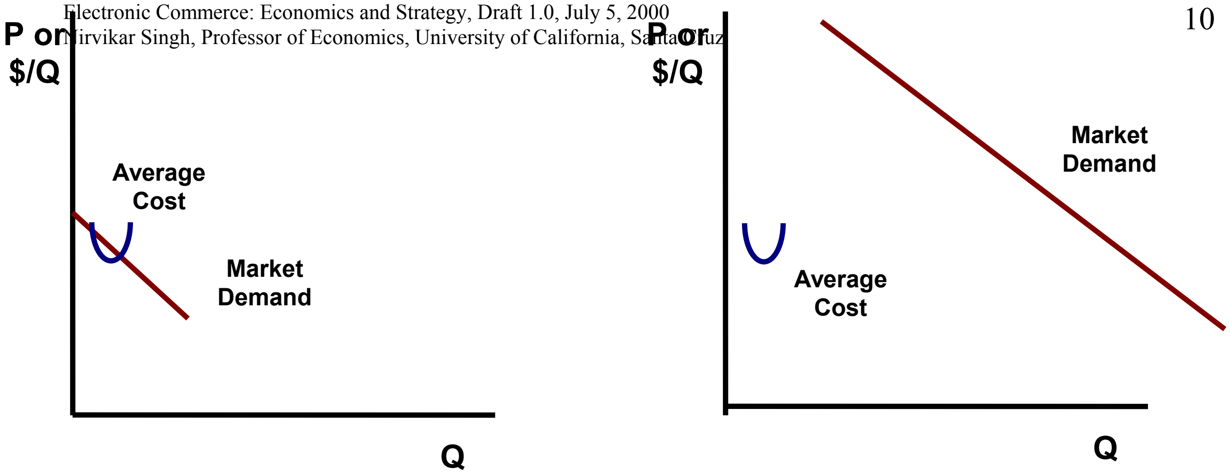


Figure 4.5b



A very similar comparison of the interaction of technology characteristics and market size can be made in the case where the technology is the same, but market demand varies. This comparison is illustrated in Figure 4.6. In the left hand panel (4.6a), the market size is small, and the firm has market power, perhaps even a monopoly. Even though increasing returns to scale apply only up to a small level of output, this level is high relative to the market size, as measured by the quantity demanded.

Figure 4.6a **Figure 4.6b**



In the right hand panel (4.6b), the market is much bigger, while the technology is unchanged, and hence the industry structure is likely to be very competitive. Of course, Figure 4.6b looks just like Figure 4.5a. Figure 4.6a looks different from Figure 4.5b only because the scale of the graph is different. In both cases, increasing returns apply at a level of output that is high relative to the market size.

We can give a different interpretation of the two panels of Figure 4.6. Think of the left hand panel as 'now', and the right hand panel as 'the future'. In this interpretation, the market is going to grow considerably. The implication of our simple analysis above is that, absent other effects, market power is going to be eroded as the market grows. New firms will enter, and the industry will become more competitive. This scenario is very important for understanding debates about the future of e-commerce in general, and of 'e-tailers' and 'dot.coms' in particular.

If one believes that a market will evolve over time as illustrated by a shift from Figure 4.6a to Figure 4.6b, then expecting that any firm will make a killing is misguided optimism. On the other hand, if the future will look like Figure 4.5b, or there is some other scenario that has a similar implication of sustained market power, then the optimists have a case. Exploring these competing scenarios will be a significant theme throughout this book.

To complete this exposition, which has informally talked about market power or its absence, we must briefly review the formal analysis of how output and prices are determined in the different cases illustrated in Figures 4.5 and 4.6. It is simplest to begin with a monopoly (single seller), where the demand for the firm's product is the overall market demand for the product. Graphically, the demand curve slopes downward, which simply means that at lower prices, the quantity demanded is higher. We can also state this idea in a different way: the marginal willingness to pay for the last unit to be sold (which is also the maximum price that can be charged for that unit) is lower at a higher quantity.

Now we make yet another simplifying assumption, relaxed in Chapter 14, that the monopoly firm must charge all its customers the same price for every unit they buy. This is reasonable if customers can easily resell to each other, and if they know that different prices are being charged to different customers. The assumption that everyone pays the same price for every unit means that, if the firm wants to increase its sales by one unit, it must lower its price for all the units it sells. For example, if it can sell a maximum of 1000 Internet service subscriptions for $20 a month, it may need to charge $19 to sell 1100. This is a consequence of the downward sloping demand curve. The additional revenue from the extra 100 subscriptions is not $1900 ($19 x 100), because now everyone is getting it for $1 less. The 1000 customers who were paying $20 are now paying $19. Hence the additional revenue from adding the extra 100 customers is $1900 minus $1000, or only $900. This is just $9 per extra customer! This value is what economists call the marginal revenue.

The general conclusion illustrated by the example is this:

> When the firm faces a downward sloping demand curve, its marginal revenue is always less than the price it charges.

This will turn out to matter a lot, because the marginal revenue is what helps determine the firm's profit-maximizing price-output choice. Ask yourself whether a firm that wants to maximize profits can do better than producing some arbitrary level of output, that it sells for the maximum price it can. If its marginal revenue from selling another unit is greater than its marginal cost of producing that unit, then doing both things adds to profit (which is just revenue minus cost). So the firm can do better by producing more. If, on the other hand, marginal revenue at that initial output is less than marginal cost, then the firm is better off reducing output by one unit! The only case where the firm can be doing the best it can, that is, maximizing profits, is at an output where marginal revenue and marginal cost are exactly equal! This output choice, Q^M, is shown in Figure 4.7.

Figure 4.7

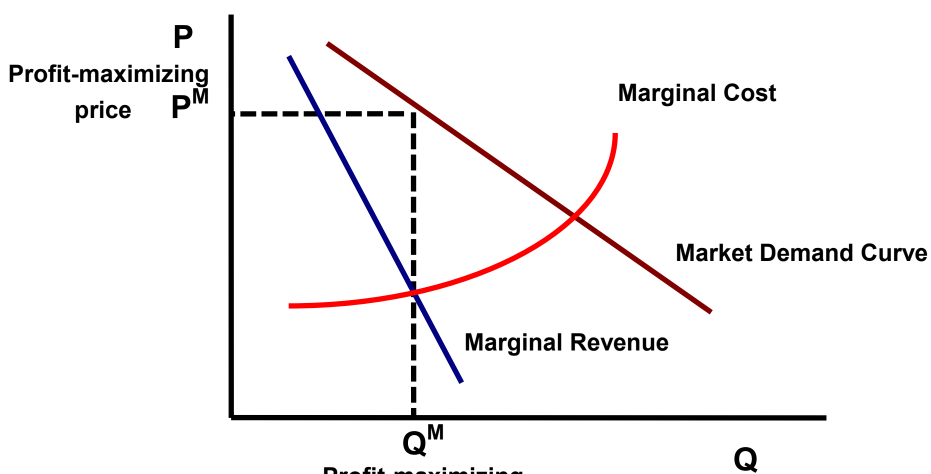


Figure 4.7 also shows the profit-maximizing price, P^M, that will be charged by the monopoly firm. This price is determined as follows. Once the profit-maximizing output is decided, the firm will charge the maximum that the marginal customer is willing to pay for the product or service. This level is determined geometrically by drawing a vertical line up from Q^M until it meets the demand curve. Note that since price is greater than marginal revenue, the profit-maximizing price is greater than the marginal cost at the profit-maximizing output. This can be thought of as a mark-up on (marginal) cost. It is

likely (though not guaranteed) that the firm's price will exceed its average cost as well. In that case, the firm must be making an **economic profit**. Recall that economic profit is different than accounting profit, because opportunity costs such as the implicit costs of capital are also subtracted from revenue.

Concept Check:
Suppose that marginal cost is constant, so that the marginal cost curve in Figure 4.7 is a horizontal line. Also suppose that there are no fixed costs. Can you see why average cost and marginal cost must be the same in this case? Try a numerical example. Suppose that the total cost of producing output Q is cQ, where c is some constant number. What is the average cost of producing Q? What is the marginal cost?

If AC = MC, can you show that the mark-up P – AC will be positive for the profit-maximizing monopoly firm?

So far, we have examined the output-price choice of a monopoly firm. In this case, the demand for the firm's product is the total demand in this market or industry. However, some of the lessons of this analysis apply to any firm that has some market power. Market power is reflected in a downward-sloping demand curve: the quantity the firm produces and sells has some impact on the price it can charge. Hence, in Figure 4.7, if the demand curve represents not the total market demand, but some fraction of it, then the rest of the analysis is exactly the same.

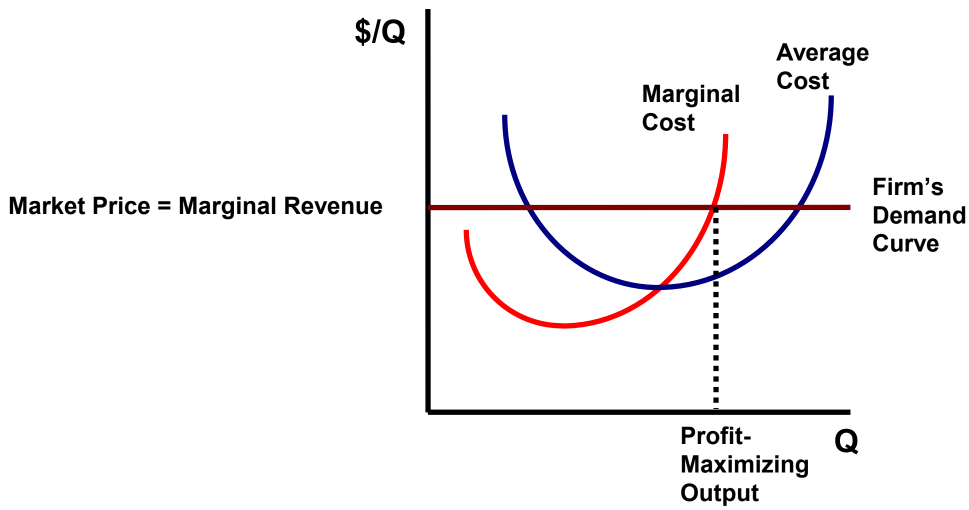
If that seems to glib, that is because it is. The full analysis is quite a bit more complicated. If there are a few firms in the industry (an oligopoly) the demand curve a single firm faces reflects what it assumes or expects the rival firms are doing. If the rivals are charging low prices, then this firm's demand curve will be further to the left, while if they are restricting output and/or trying to charge high prices, then this firm's demand curve will be further to the right. So working out what happens in the market requires looking at all the firms together. We postpone this analysis to later in this section.

Now we look at the remaining possibility, where the firms in an industry have no market power. This is an extreme case of the kind of situation illustrated in Figure 4.5a and Figure 4.6b. Now the market demand curve does not directly impact the firm's choices. It does matter, of course, since market supply and demand determine the market price for the product. But it is the market price that determines what the individual firm does, not the whole demand curve.

One way to think of this is that the individual firm's demand curve is simply a horizontal line, at a level that reflects the market price. The firm's output choices have no effect on this market price. Unlike the cases of market power, the firm's choice is no longer a simultaneous determination of the profit-maximizing output and price: the firm with no market power takes the price it can charge as given by the market price.

If the demand curve facing the firm is horizontal, then it can sell whatever it likes at the going price. Note that this is not strictly true. If the firm starts getting big, its output will start to 'move the market'. We are assuming that the firm is so small that it can not reach such outputs at a reasonable cost: look at the average cost curves in Figures 4.5a and 4.6b. Since we do not need to be concerned about the market demand curve, it makes sense not to use the horizontal scale of those figures, but instead to stretch out the horizontal axis, as we do in Figure 4.8.

Figure 4.8



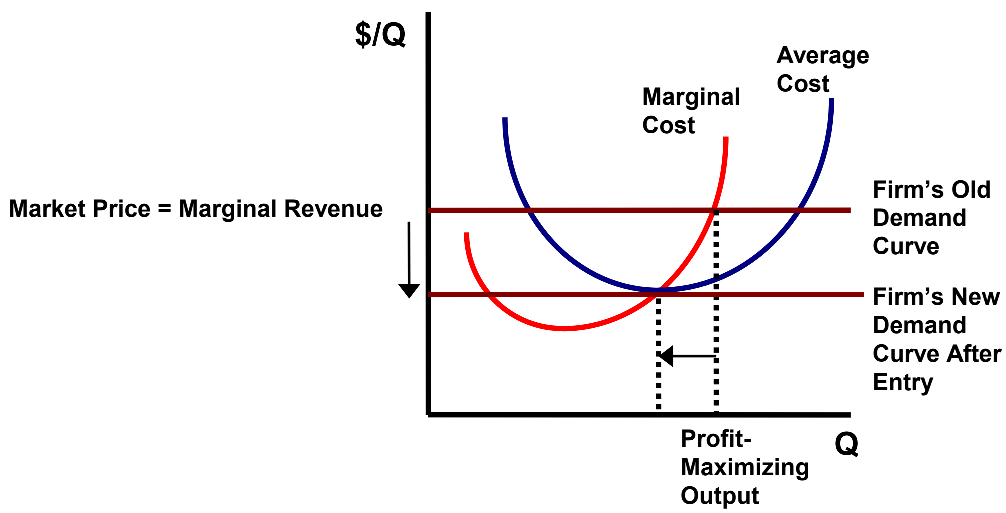
In Figure 4.8, the horizontal line is the firm's demand curve, and its height is determined by the market price, whose determination we are not analyzing for now. Since the firm can sell whatever it wants (in the range of output shown) at the going price, the extra revenue it receives for an extra unit produced and sold is just the market price. Hence the price and marginal revenue coincide (P = MR) in the case of the firm with no market power. This is different from our market-power case, where P > MR.

However, our previous argument for determining the profit-maximizing output still holds: it is the output level at which MR = MC. In this case, we have P = MR = MC at the profit-maximizing output level, as shown in Figure 4.8. Notice that we have drawn the average cost curve in Figure 4.8. It was not in Figure 4.7 only because we chose to leave it out to keep the picture simple, and because we wanted to focus on the concept of marginal revenue. In Figure 4.8, since the marginal revenue curve coincides with the firm's demand curve, we have put back the firm's average cost curve. You can see that, at the profit-maximizing output level, the price the firm sells at is less than its average cost (P >AC). The difference, P − AC, is the firm's economic profit per unit of output. If we multiply it by the firm's total quantity of output or sales (the two are identical in this simple case), then we get that the firm is making that level of economic profit.

Positive economic profits represent returns over and above what the resources used in the firm's production could earn elsewhere in the economy. That is good for the owners of the firm. If the firm is a public company, with shares traded on the stock market, the owners of the firm are the shareholders. The value of the firm's shares will reflect the economic profits that the shareholders expect the firm to make over time. We will discuss this issue in greater detail in Section 4.5, since it is at the heart of the debates about the high market values of dot.coms, or e-commerce or Internet firms in general. Next, we examine how a situation such as that in Figure 4.8 might evolve – we argue that such a situation is unlikely to persist for long.

Entry and Exit Figure 4.9 repeats Figure 4.8, but also shows what will happen over time to the firm. We have noted that the firm in Figure 4.8 is making positive economic profits. These represent above-average returns to the capital and skills employed in that industry. It could be that the firm has some special advantage over other firms in the industry, that gives it lower costs. The most obvious example would be some intellectual property (say, software that governs its production processes) that other firms can not imitate, either because it is a trade secret, or because it is patented (recall Chapter 3's discussion). For now, let us rule out this case, and assume that the firm has no such advantage. In other words, new firms can enter the industry and do exactly what this firm is doing. That has been the experience of many e-tailers: they are all quite new, they have similar technology and sell similar or even identical products. Thus, in this story, branding, reputation and product differentiation are all absent.

Figure 4.9



If there is nothing special about this firm or any other firm in the industry, and there is no legal or regulatory barrier, then any other firm should be able to set up shop – on the web in the e-tailer example – and do whatever the existing firms are doing. In that case, the market price will not stay the same, but will fall. We are assuming that this entry takes place much faster than the market can grow, which seems to be a realistic assumption. Only when the market price falls to the level at which the e-tailer is making no economic profit (it still might be making profits in the accounting sense, representing a return on the equity invested in the firm) will there be no incentive for new firms to enter.

Since the firm with no market power maximizes profit by producing the output at which price equals marginal cost (P = MC), and since no economic profit means that price and average cost are the same (P = AC), it must be true that entry will stop when marginal and average cost are equal, as shown in Figure 4.9. Mathematical relationships ensure that this must be at the bottom of the average cost curve's U-shape. More important than the geometry, however, is the conclusion that if firms do not have a special advantage, entry will drive down prices until returns are just what could be earned anywhere else.

This lesson might not seem to apply to many 'dot.coms' or other Internet firms, that are spending heavily to build brands, and differentiate their products and services. If one of these firms is successful, after all, then it will have market power, and the demand curve it faces will be downward sloping, like that in Figure 4.7. However, even in this case, entry will push the firm's demand curve down and to the left (it can sell less at any price, or charge less for any level of output it wishes to sell). Unless there is some other force at work, the firm's economic profit will approach zero. Much of business strategy, then, consists of how to avoid this fate! We discuss this more in future chapters.

We conclude this part with some observations on exit. Suppose, for some reason (optimistic entrepreneurs, or optimistic investors!), that there are 'too many' firms in the industry, so that the market price is below the bottom of the average cost U-shape. In that case, every firm will be making economic losses. Now, these may still represent accounting profits, but the owners of the firm will not be getting a return as high as they could elsewhere (by definition of economic profits), and so this situation can not persist.

Eventually, firms will leave the industry. They may not do so immediately, because if they have sunk some capital into the industry, then exiting will not recover that part of their costs. Hence firms can go on producing for some time at a loss. Finally, however, there will come a point where, unless something else happens, firms must start to exit. When they do, then prices will rise, and the remaining firms in the industry will start to earn a normal rate of return (zero economic profit). A process such as this has already begun with the 'dot.coms', as of this writing (June, 2000).

Strategic Behavior Earlier, we put aside the fuller analysis of what happens when there are just a few firms in an industry, so each has market power, but at the same time they have to be conscious of what their rivals are doing or might do, as they make their own pricing and output decisions. This kind of situation is actually pervasive, and the monopoly and pure competitive (zero market power) cases are merely idealizations at either end of the continuum of possibilities of strategic interaction.

Here we illustrate some of the simplest ideas for analyzing strategic behavior. At the same time, they are emblematic of general and powerful approaches to the problem. Let us assume that there are only two firms in the industry (online booksellers, say). Let us also suppose that, when a firm is deciding on its price and output combination, it assume that its rival will hold its output constant, perhaps because it has already committed itself through contracts to employees, leases for equipment, and so on. Then the firm's demand curve is whatever residual demand there is at any price, subtracting off the rival's output. However, the firm doesn't know exactly what the rival is going to produce. If it plans its profit-maximizing response for each possible output level of the rival, this will trace out a curve showing how it responds to different output levels of the rival.

Figure 4.10a

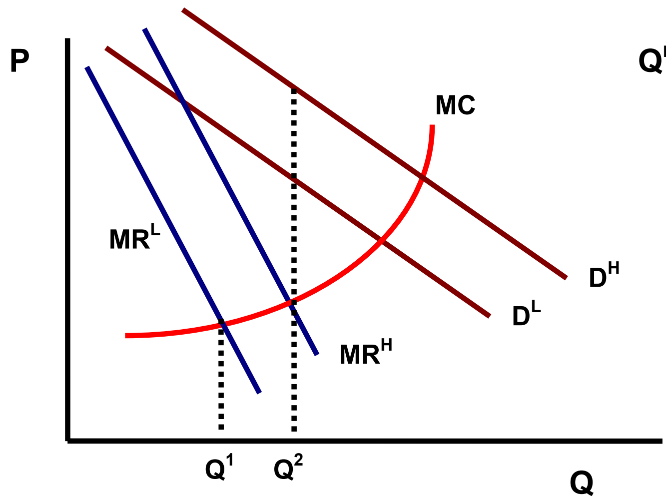


Figure 4.10b

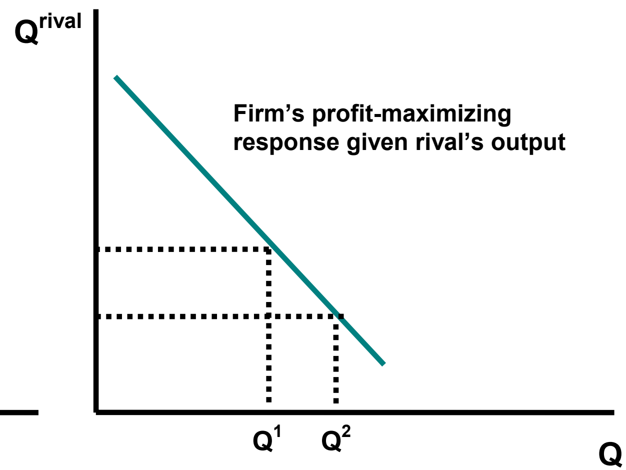
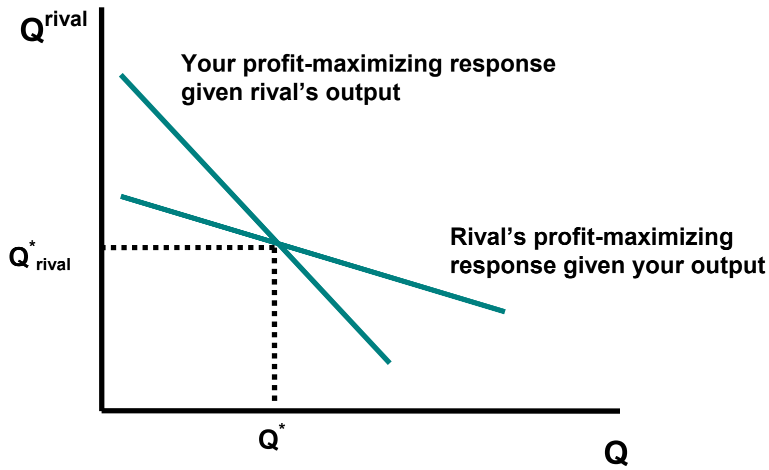


Figure 4.10a is similar to Figure 4.7, except it shows two possible demand curves for the firm, and the two corresponding marginal revenue curves. D^L and MR^L are what is relevant if the rival has one particular output, while D^H and MR^H characterize the situation if the rival's output is another, *lower* level. These are just two possible guesses about what the rival will do. The firm's profit maximizing response in each case is given by Q^1 and Q^2, respectively, as shown in Figure 4.10a. If we plot the different profit-maximizing responses of the firm against all possible guesses about the output level of the rival, we get the line shown in Figure 4.10b. This is a very simple form of strategic behavior, but it illustrates the general idea that what is best for you to do depends on what your rival is likely to do.

What will your rival actually do? If it can anticipate your responses, it can choose the point on the response line in Figure 4.10b that maximizes its own profit. In that case, the rival is the leader, and you are the follower. If both firms are fairly similar, however, there is no reason for this asymmetry in anticipation and behavior. In that case, your rival will have the same guesswork to do, and will have the same kind of profit-maximizing responses to all your possible output levels. Its profit-maximizing response curve will also be downward-sloping, but it will be drawn from the rival's perspective. We can put both curves on the same diagram, as shown in Figure 4.11. The rival's curve is flatter only because it is drawn from the rival's 'sideways' perspective.

Figure 4.11



The two profit-maximizing response curves are shown intersecting. This intersection point determines your output, Q*, and your rival's output, Q^*_{rival}. At Q*, you are maximizing your profit, given that your rival produces Q^*_{rival}. And your rival is maximizing its profit by producing Q^*_{rival}, given that you are producing Q*. Thus each of you is doing the best you can, given what the other is doing. Neither of you has an incentive (unless something changes, or you anticipate a change) to produce either more or less. This situation is called a **Nash Equilibrium**, after the Nobel-prize winning mathematician, John Nash, who formulated this idea. The concept of Nash Equilibrium is a very general and powerful one, but Figure 4.11 gives one of its simplest illustrations.

Concept Check:
Think of an argument to justify the statement that the intersection of the two curves in Figure 4.11 is the only Nash Equilibrium in that situation. (Hint: start with any other point in the figure, then argue that one or both firms can do better by changing its profit-maximizing output, given the other firm's output level.

Figure 4.11 also illustrates another important aspect of strategic behavior. In the figure, each of the response curves is downward-sloping. This reflects the fact that, if your rival is producing more, your marginal revenue curve is lower, and so your profit-maximizing output as a response to your rival's output is also lower. This was the intuition illustrated in Figure 4.10a. We have a name for a situation where the level of your strategic variable (here, output) is adjusted downward as the level of your rival's strategic variable increases. The strategic variables here are called **strategic substitutes**: one firm's output gets *substituted* for the other firm's output.

The polar case to that of strategic substitutes is one where the firms' profit-maximizing response curves are upward-sloping. How can this happen? Suppose that the two firms' products are somewhat different, so that they can get away with charging different prices. In the example so far, we assumed that the firms decided on their output levels, and the price was then simply the market price, at which all this output could be sold. Now let us suppose that firms set their prices, see how much of their product will be demanded at those prices, and then produce to order. (Whether this assumption is more realistic than the previous one depends on a host of factors, that we can not elaborate on here.) How much a firm can sell, and its profit-maximizing price, will depend on the price that its rival charges. Here, however, if your rival charges a higher price, your marginal revenue curve will be higher, and you can charge a higher price. In the earlier situation, a higher output by the rival gave you 'less room' in the market. Here, a higher *price* by your rival gives you *more* room, and you can take advantage of that with a higher price.

Figure 4.12

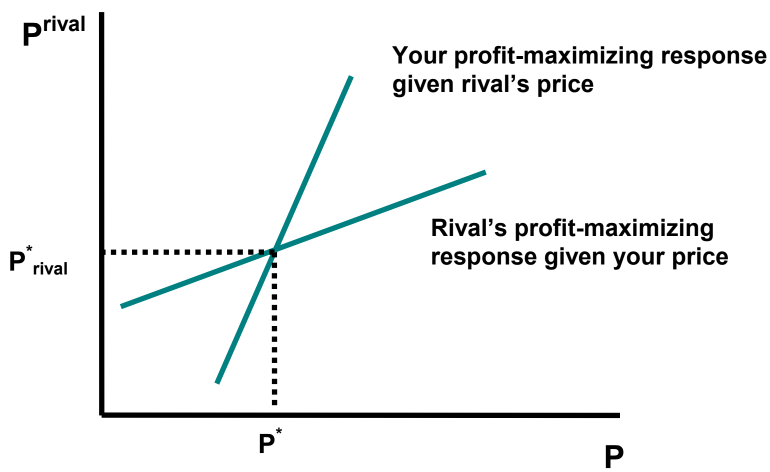


Figure 4.12 illustrates the situation. Note that the figure shows that you will not raise your price as much as your rival, in response to a price increase by your rival. This seems plausible, and is true under reasonable assumptions about the marginal revenue and marginal cost curves. The intersection of the two new profit-maximizing response curves is now also a Nash Equilibrium, but a different kind than earlier, because the strategic variables (and the 'rules of the game' in general) are different. You can check that it is a Nash Equilibrium, and the only one in the figure. In this case, since the strategic variables, the prices, tend to move together, they are called **strategic complements**.

The distinction between situations of strategic substitutes and strategic complements is an extremely valuable one in practice, and it goes well beyond the simple examples here. The basic idea involves deciding whether, when you change the level of one of your strategic variables, your rival will respond aggressively or in an accommodating manner. Besides output and price levels, all real-world firms (not just in e-commerce) have to decide on advertising and marketing expenditures, product characteristics, capacities, and a host of other strategic variables. Being able to predict the *direction* of a rival's response is an important starting point for strategic analysis.

We close this discussion of strategic behavior with an example where there are only two levels of the strategic variable. In this example, two music companies can choose whether to have liberal or strict copyright enforcement policies. In the example, Atlantic and Sony are two music companies, each with its own set of contracted artists.

Thus their products are not perfect substitutes, but they are close enough so that greater ease of access to the music may cause customers to switch. Tables 4.1a and 4.1b show two different possibilities. While profits were in the background in Figures 4.11 and 4.12, here they are explicitly given: in each cell of each table, the first number is Atlantic's profit in millions of dollars, and the second number is Sony's in the same units.

Table 4.1a

Sony / ***Atlantic***	**Liberal**	**Strict**
Liberal	(90, 90)	(110, 80)
Strict	(80, 110)	(100, 100)

Table 4.1b

Sony / ***Atlantic***	**Liberal**	**Strict**
Liberal	(80, 80)	(110, 90)
Strict	(90, 110)	(100, 100)

In Table 4.1a, the strategic variable pair (Liberal, Liberal) is the only Nash Equilibrium. Given that Sony is liberal, Atlantic has a higher profit from being liberal than from being strict (90 versus 80 million). Since the situation is symmetric, the same holds for Sony. (Strict, Strict) is not a Nash Equilibrium, because each company will have an incentive to be liberal, as long as the other remains strict: profits increase from 100 to 110 million as a result. (Strict, Liberal) is also not a Nash Equilibrium, since Atlantic has an incentive to shift to being liberal (why?). By symmetry, (Liberal, Strict) is also not a Nash equilibrium. Thus we have established that (Liberal, Liberal) is the only Nash Equilibrium.

In Table 4.1b, (Strict, Strict) is still not a Nash Equilibrium, using the same argument as before. But now we have switched some profit figures, and (Liberal, Liberal) is also not a Nash Equilibrium: each music company will want to switch to being strict. However, now (Strict, Liberal) is a Nash Equilibrium! By symmetry, so is (Liberal, Strict). This second table therefore illustrates two new ideas:

1. There can be more than one Nash Equilibrium in a given strategic situation.
2. Nash Equilibria can involve asymmetric behavior, even when the strategic situation is symmetric.

Concept Check:
Think of an argument to justify the statement that (Strict, Liberal) is a Nash Equilibrium in the situation of Table 4.1b. (Hint: start with that strategic pair, and explain why neither music company will want to switch).

4.3 What Households Do

If firms produce and sell, households buy and consume. Economics principles texts focus on explaining how households' purchasing and consumption decisions underlie the market demand curves that help determine market prices. We will review these ideas, and also extend them to cover three aspects of household decision making that are often swept aside, but that are very important in analyzing the economics of e-commerce. These are: the availability of information, the use of time, and allocation of savings. Much of the potential advantage of online shopping over traditional approaches lies in more efficient gathering of information and use of time. The importance of the allocation of savings has a more complex motivation. The allocation of savings is mostly about the choice of financial assets (unless you buy diamonds, gold and art). Financial assets are claims on future purchasing power, and ownership and transfers of these assets are all simply information that can be stored, processed and transmitted in digital form. This makes online personal management of financial assets a potentially attractive option for many households.

Households are, in their own way, organizations just like firms. Just as we have put aside much of the internal workings of firms in Section 4.2, in order to draw out some basic principles of how firms behave, we do the same for households. More and more households are single individuals, in any case, so treating the household as an individual is perfectly accurate for them! Therefore, let us begin with the fundamental motive of an individual, which is to maximize his or her satisfaction, happiness or 'utility'. The last of these terms is the economist's preferred jargon. All kind of factors influence a person's level of satisfaction or utility, but to start with, let us suppose that it depends on the consumption of quantities of a set of products and services. To keep matters really simple, we assume that this set includes just pizza and music, and that these categories are perfectly homogeneous (all pizza is the same, all music is the same).

Now suppose that a typical person, say, Shawn, has a given weekly budget to spend on buying pizza and music (this is before Shawn's namesake invents Napster!), and he can buy as much as he wants at the going prices, as long as he stays within his budget. You may recall from your economics principles course that Shawn will maximize his weekly utility (satisfaction) by allocating his dollars so that

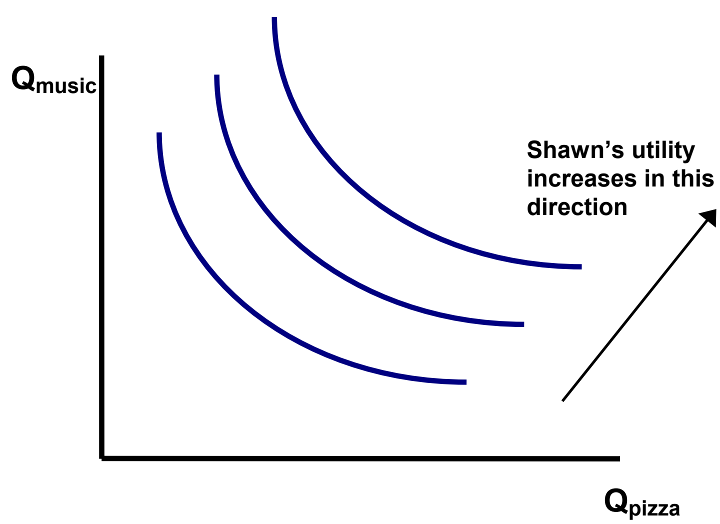
$$\frac{\text{Marginal utility of pizza}}{\text{Price of pizza}} = \frac{\text{Marginal utility of music}}{\text{Price of music}}$$

The logic of this equation is a familiar one: if it does not hold, then Shawn can increase his utility by reallocating his dollars. For example, suppose that the left hand side (MU_{Pizza}/P_{Pizza}) is greater than the right hand side. Then Shawn is getting a bigger 'bang for the buck' with the last dollar spent on pizza than the last dollar spent on music. He can increase his utility be shifting a dollar from buying music to buying pizza. To keep this simple, we are assuming that he can make these shifts in small increments. Of course that is not always possible: you can buy a slice of pizza, but not one bite! Changing our assumption does not change the central message, though.

Concept Check:
For the argument about reallocating dollars to have an ending where Shawn is still buying some music, it helps to assume that the marginal utilities of consuming pizza and music are diminishing: they go down as consumption of the product goes up. Why does this help? Is it a plausible assumption?

Now we are going to introduce a graphical tool to look at Shawn's choice in a different way. This graphical tool is the **indifference curve**. An indifference curve, in this example, shows all the combinations of pizza and music that give Shawn exactly the same level of utility. For each such level of utility, there is one indifference curve. Therefore, we could draw an infinite number of indifference curves, but we only show a few in Figure 4.13, to illustrate their typical shape.

Figure 4.13



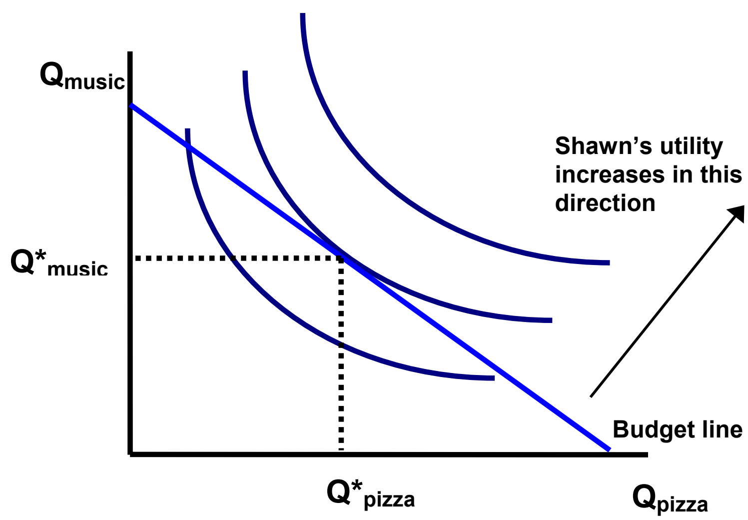
Shawn would like to move his consumption up and to the right as much as possible, since we are assuming that consuming more always makes him happier. He is constrained only by his budget. Say his budget is B dollars. Then his spending must obey the rule which describes this 'budget constraint':

$P_{pizza}Q_{pizza} + P_{music}Q_{music} = B$.

This is the equation of a straight line with a negative slope. P_{pizza}, P_{music} and B are the constants, while Q_{pizza} and Q_{music} are the variables.

Figure 4.14 repeats Figure 4.13, but with the budget constraint added. The middle indifference curve happens to be the highest one that Shawn can get to, and the point where it touches the budget line is the combination of quantities of pizza and music that makes Shawn happiest – maximizes his utility – that week. It can be shown mathematically that this combination (Q^*_{pizza}, Q^*_{music}) also satisfies the requirement that the marginal utility per dollar must be the same for pizza and music.

Figure 4.14



Using either the marginal utility per dollar approach, or the indifference curve approach, we can argue that if the price of one of the products, say pizza, falls, Shawn will increase his purchases of pizza. Thus Shawn has a downward-sloping demand curve for pizza. Combining all the individual demands for pizza in the neighborhood where Shawn lives, we will get a downward-sloping market demand curve for pizza. Note that, as the price of pizza falls, and Shawn buys more pizza, as long as nothing else has changed, he must be buying less music. Shawn substitutes pizza for music, and a fall in the price of pizza therefore shifts Shawn's demand curve for music to the left. You will have dealt with such concepts in your economic principles course.

Now we are ready to tackle some of the issues that relate more closely to online shopping. The story so far has assumed that Shawn knows the prices of pizza and music, and that time is not a relevant factor, either in shopping for the two products, or in consuming them. Online shopping is of interest precisely because these assumptions often do not hold in practice: traditional shopping takes time. We will begin with some more basic aspects of time, particularly its role in consumption, and then look at time in shopping, along with the role of information.

Time, Work and Consumption Leisure time typically provides us with utility (satisfaction). We can incorporate this into Shawn's decision making, while simultaneously solving the mystery of where his weekly budget comes from. Suppose that Shawn can work as many hours a week as he wants, at a given wage (no overtime). If the wage is W dollars per hour, then Shawn can earn a maximum of 168W dollars in a week. Of course, Shawn will not want to go without sleep (unless he can take naps on the job, that is what working 168 hours a week involves). For each hour he chooses not to work, Shawn gives up $W. In doing so, it is as if Shawn is purchasing his leisure time. His wage rate is therefore the price of his leisure!

Now we can describe the condition under which Shawn's choice of pizza, music and leisure will be utility-maximizing. For utility maximizing, the marginal utility per dollar must still be the same, so we have

$$\frac{\text{Marginal utility of pizza}}{\text{Price of pizza}} = \frac{\text{Marginal utility of music}}{\text{Price of music}} = \frac{\text{Marginal utility of leisure}}{\text{Price of leisure (wage)}}$$

Since we now have three things that Shawn is consuming, drawing the analogue of Figure 4.14 would involve one more dimension. To avoid that, instead suppose that Shawn works at a pizza place, where he gets free pizza, so that all his budget goes on music (we will assume that his needs for clothes, shelter, and so on are taken care of some other way!). His budget constraint is now:

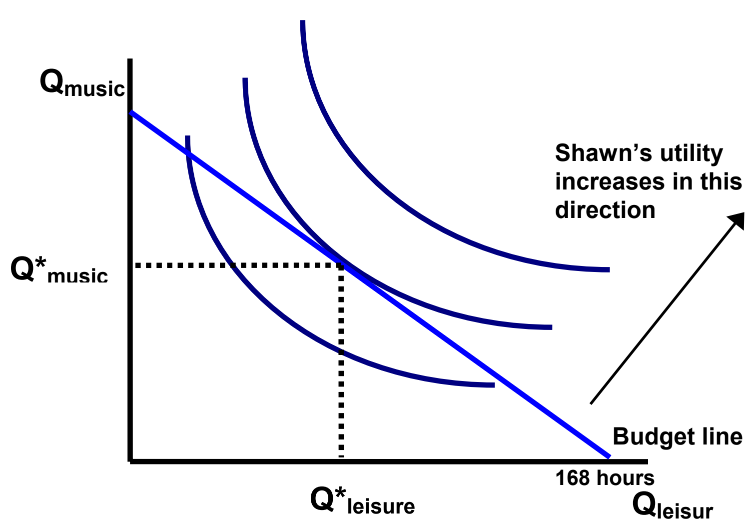
$P_{music}Q_{music} = B = W(168 - Q_{leisure})$.

$(168 - Q_{leisure})$ is how many hours a week that Shawn works, and multiplying it by the wage gives his budget. We can rearrange this budget equation to emphasize the idea that Shawn is implicitly purchasing leisure:

$P_{music}Q_{music} + WQ_{leisure} = W168$.

In any case, Shawn's choice can now be shown in two dimensions, as in Figure 4.15.

Figure 4.15



What does Shawn do with his leisure time? It could be that he sleeps, runs and watches TV, and these activities involve no expenditure. However, it is plausible that Shawn also listens to the music that he buys. He might derive satisfaction from collecting the music, but listening is likely to be an important source of satisfaction. If Shawn can not listen to music at work, then he needs leisure time to enjoy the music he buys. The budget constraint in Figure 4.15 is drawn to show the tradeoff between the purchase of music and the 'purchase' of leisure (by working less than 168 hours a week). But Shawn's utility depends on his consumption of music, which depends on both his purchase of music and how much leisure time he has. This is not too bad, because ultimately, his utility just depends on the two things he purchases: music and leisure time. Leisure time provides some direct utility, and also acts as an 'input', together with his music purchases, in Shawn's consumption of music, which gives him utility as well.

If Shawn has several leisure activities that involve spending money as well as time, he can make his decisions as follows. For each possible choice of hours worked, he has a budget of money and a 'budget' of leisure time. Shawn can figure out his hypothetical allocation of time and money among his leisure activities based on what they cost in terms of time and money, and his two budgets. If it seems that he has too much money (based on the particular work option) relative to time, he can work less – like the doctor who takes off two afternoons a week to play golf. If it seems money is relatively scarce, he can plan to work more hours. Of course, in the real world we use rules of thumb rather than such elaborate calculations, but the general effect is the same. Often, though, we do not have this kind of flexibility, especially in the short run. The number of hours we work is not so freely adjustable.

Time and Shopping We have introduced time as something of value. You can use time to earn money, to enjoy leisure directly (without buying anything), and to enjoy the consumption of things you buy. Time is also used up in the process of purchasing things. Shopping may itself be an enjoyable activity, just as much as listening to music, but it is often simply a chore. How does time spent in shopping change Shawn's decision making? Consider again the case where Shawn has to decide how much work and how much to enjoy leisure and music (Figure 4.15). His budget constraint when he did not have to spend time shopping was:

$P_{music}Q_{music} = W(168 - Q_{leisure})$.

Now suppose that for each unit of music he purchases, he has to spend t hours shopping. This may be time spent in going to music stores, finding what he wants, making choices, and so on. We neglect the money costs of shopping, such as travel costs, but those may be easily added in as well. Again, shopping is often an enjoyable activity, but we assume for now that it simply reduces the time available for enjoyable leisure pursuits. Thus Shawn's budget constraint is now:

$P_{music}Q_{music} = W(168 - Q_{leisure} - tQ_{music})$.

We have subtracted from 168 (the number of hours in a week) not only Shawn's leisure time, but his shopping time, to get his net time working and earning. We can rearrange this equation as

$P_{music}Q_{music} + WtQ_{music} = W(168 - Q_{leisure})$.

Comparing this with the no-shopping time case, it is therefore as if Shawn's total price paid for music is $P_{music} + Wt$. The second term represents the time spent per unit purchased, valued in terms of the leisure foregone, or the earnings opportunity sacrificed.

It may seem unrealistic for the time taken in shopping to be proportional to the amount purchased. Some costs of shopping are fixed costs that are the same no matter how much is purchased. If we assume that there are *only* such fixed costs, denoted by T, then Shawn's budget constraint is:

$P_{music}Q_{music} = W(168 - Q_{leisure} - T)$.

We can rearrange this as:

$P_{music}Q_{music} + WQ_{leisure} = W(168 - T)$.

Now the price of music is no higher than in the no-shopping-time case, but Shawn no longer has 168 hours to allocate between work and leisure (unless he only works and never shops).

The two cases of shopping cost are shown in Figure 4.16. Figure 4.16a shows the case of a per-unit-purchased cost of shopping, while Figure 4.16b shows the case of a fixed time cost of shopping. In each case, the no-shopping-cost budget line is also shown. The magnitude of the shopping costs is exaggerated to show the effects dramatically.

Figure 4.16a

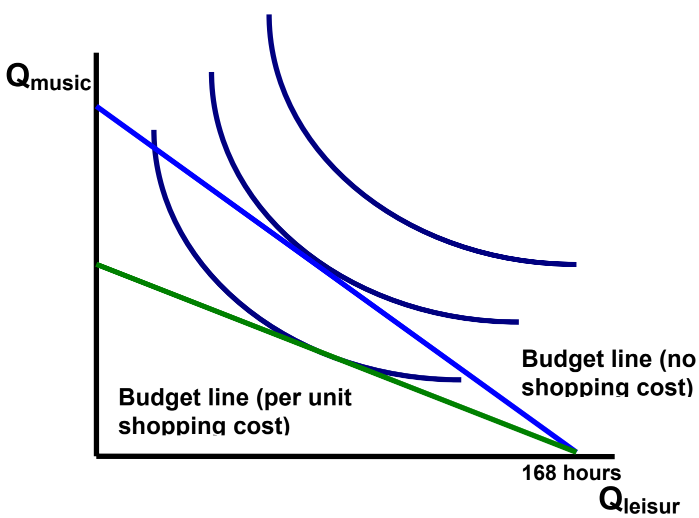
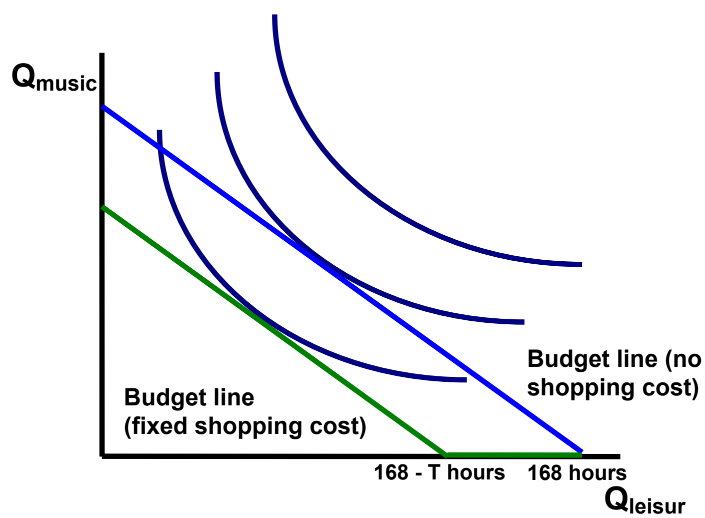


Figure 4.16b



The purpose of this rather long discussion of shopping costs has been because one of the 'value propositions' of online retailing has been precisely that it reduces shopping costs. In this respect, it is not entirely new, of course. Catalogue, mail-order and telephone shopping are precursors of electronic retailing. However, the scope of attempts in electronic retailing outweighs the traditional catalogue business. Groceries and pharmaceuticals are two important categories where print catalogues and telephone orders would be just too costly or otherwise impractical. Of course there are other differences in online retailing versus traditional approaches, and we shall explore these throughout the book. Here we may note one point that our analysis brings out.

The discussion of costs of shopping assumed that these costs were well-measured by the leisure forgone, and that leisure, in turn, was valued at the wages forgone. However, peoples' actual behavior suggests many do not value the time spent in shopping in this way. They may enjoy it, or they may not have alternative uses for time that are highly valued. They may also be creatures of habit, and not have done the kinds of detailed calculations required to realize how valuable their shopping time is. Finally, people may organize their lives to reduce shopping time to a minimum (stopping by the store on the way home from work, for example), so that the actual numbers that replace 't' and 'T' in our discussion may be really quite small. All of these factors will work against electronic retailers, though habits may change over time, and, as society becomes richer, the value of that time spent in shopping may increase.

Time and Information When we discussed Shawn going shopping for music, we explained the costs in rather sweeping terms, including possible time taken in visiting stores, and the time spent in making choices. It is now useful to separate out the information gathering part of these shopping costs, and look at the process of information gathering more closely. Again, this is something that economic principles courses may underplay, but it lies at the heart of the Internet and online commerce. Therefore we must face it head on.

Information gathering is often a significant component of shopping time. In a weekly trip to the same grocery store, with the store's sale flyer in hand, the shopping time is almost all purely routine: driving to the store, walking down familiar aisles and loading the cart, and so on. On the other hand, shopping for things like cars, stereos and furniture involves inspecting the products, trying them out, comparing features, asking questions from the sales people, and asking friends about their experiences with similar products. In other words, there is substantial time (and maybe monetary costs as well) devoted to information gathering as part of the purchase decision. None of this is really considered in the beginning textbook model, where Shawn's budget constraint for pizza and music (illustrated in Figure 4.14) assumes that he knows exactly what he wants, as well as the prices of the two products.

Part of the problem is that, in the real world, products and services have a great deal of variety, and their characteristics may not be known without investing the time to acquire the information. This is not captured in the textbook model of homogeneous pizza and music. Even a simple thing like pizza may have too much cheese, or not enough, or the wrong kind, and so on. Sometimes, you just have to try a product or service to be able to judge it, but even there, some prior research may help you avoid a mistake: you can read the ingredient list or ask friends who have eaten that kind of pizza before. Since we have not formally discussed the case of product variety or product differentiation, we postpone these kinds of information issues till future chapters.

In addition to gathering information about product characteristics, an individual can also spend time finding out which seller charges the lowest price. If the product is homogeneous, it would seem reasonable that the price should be the same for every seller. That was implicit, for example, in the earlier discussion of what happens when firms do not have market power. However, that discussion assumed that buyers would costlessly learn the range of prices being charged, and go only to the lowest-priced seller. Now let us do away with this assumption.

There are many ways of thinking about how individuals traditionally gather price information. They may search advertisements, call up or visit different stores, or ask friends. Analyzing these methods in detail can be quite complicated, but we can give a simplified treatment. We begin with the case illustrated in Figure 4.15, where Shawn works to earn money to buy music. Suppose that the price that Shawn can get for music is lower the more time he devotes to searching for deals, sales or special offers. If S is Shawn's search time, his budget constraint is:

$P_{music}(S)\ Q_{music} = W(168 - Q_{leisure} - S).$

The 'S' on the right hand side appears in the same place as 'T' in the fixed shopping time example. Now however, Shawn can choose 'S', and a greater time spent searching leads to a lower price paid, so 'S' provides a benefit as well as a cost. The choice of purchases of music and how much leisure to enjoy is still subject to the 'bang for the buck' rule:

$$\frac{\text{Marginal utility of music}}{\text{Price of music}} = \frac{\text{Marginal utility of leisure}}{\text{Price of leisure (wage)}}$$

However, the price of music in this equation depends on another balancing act by Shawn. His marginal cost from searching another hour is always W dollars, given our assumptions about his work options. His marginal benefit of searching is given by $-\Delta P_{music}(S)\ Q_{music}$, where the symbol 'Δ' stands for 'change in'. Since searching an hour more reduces the price paid, $\Delta P_{music}(S)$ is negative. The marginal benefit per unit of music purchased is this quantity multiplied by - 1, and the marginal benefit overall is obtained by multiplying the per unit marginal benefit by the quantity purchased. Shawn's rule for how much to search is therefore to continue until the marginal benefit from searching equals the marginal cost, or

$-\Delta P_{music}(S)\ Q_{music} = W.$

This suggests that if Shawn is paid a higher wage, he will search less. This certainly is reasonable, and accords well with what we seem to observe.

Since 'S' is a third choice variable for Shawn, and the price paid depends on how much he searches, his decision-making is not easy to illustrate in a figure like Figure 4.16. However, we can make a different assumption about search that can be pictured, and also lets us compare the behavior of different kinds of individuals. Therefore, now suppose that the search cost is a fixed time 'S'. If this fixed time is spent in searching, Shawn will find where he can buy music at a lower price than if he does not search at all. In Figure 4.17a, we show the situation facing Shawn. There are two possible budget lines, depending on whether he searches or not. In the figure, Shawn is better off not searching. The best he can do is point A, on his 'no search' budget line. No point on the 'search' budget line gives him as high a utility as he gets at point A – every point on the 'search' budget line is on a lower indifference curve for Shawn. As a result of not searching, Shawn pays a higher price than otherwise.

Figure 4.17a

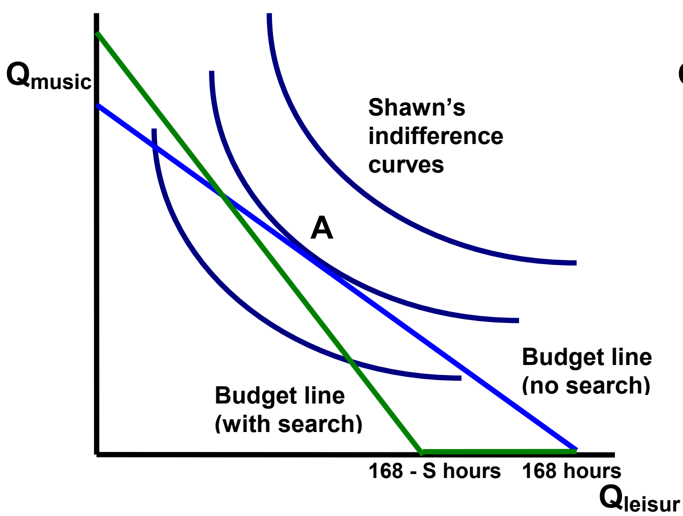
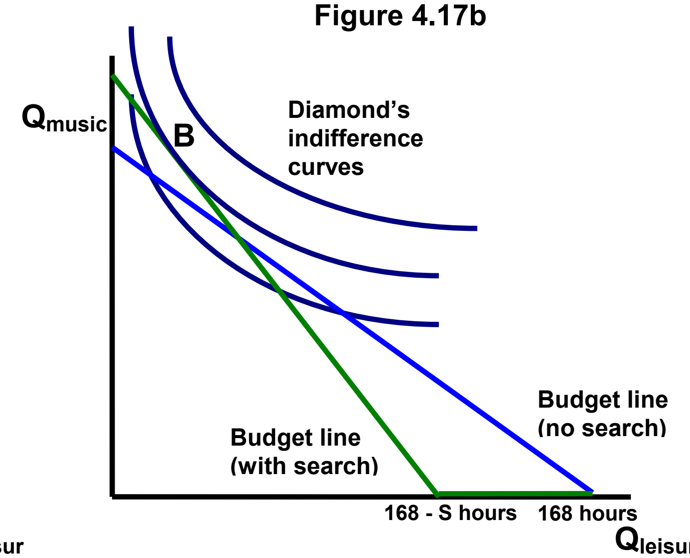


Figure 4.17b



By now you are better acquainted with Shawn than you probably wanted to be, so let us switch to his friend, Diamond. Diamond faces exactly the same situation as Shawn, but her indifference curves are rather different, as you can see from Figure 4.17b. Diamond likes music relatively more than Shawn does. If search were not possible, Diamond would choose to work more hours than Shawn, and spend that extra money on music. If she can search for S hours and get a lower price on music as a result, she does even better. At point B on the 'search' budget line, she is on a higher indifference curve than any she could reach on the 'no search' budget line.

Our conclusion is that Diamond searches and gets a lower price, and Shawn does not search, and pays a higher price. In the illustration, this is because Diamond gets a greater benefit from searching. We could get the same kind of outcome if Shawn and Diamond had the same preferences, but Diamond was better at searching than Shawn, and could do it in considerably fewer than S hours. In any case, differences in customer profiles lead to different behavior. Identifying these kinds of differences is part of the task of marketing, which we consider in Chapter 15. Pricing is also dependent on these differences. Music sellers may (and do) charge different prices for exactly the same product, if there are consumers like both Shawn and Diamond around.

We can also use Figure 4.17 to illustrate one possible consequence of online selling. If selling online reduces the time required for searching, then it just means that Diamond searches for less time and still gets the low price for music, and she can be better off (to the northeast of B, on a new, higher budget line). If the time required to search falls enough, however, Shawn's behavior changes. If his 'search' budget line shifts out enough, he now finds it worthwhile to search, and he, too, pays the lower price. If no one is buying from high-priced sellers, then they may cut their prices. This can reduce the benefit of search so that music buyers like Shawn once again do not search.

The predictions of this simple model are straightforward ones: online commerce reduces price dispersion as well as average prices. It does not follow that price dispersion is eliminated. This could only happen if search costs are zero. The real world has several complicating factors: search is more complex, it involves sequential decision-making (not just whether to search or not, but when to stop, based on what you have observed so far), and it is multi-dimensional (not just price, but service, quality, reputation and so on). Buyers may also still be mastering the art of shopping online, whereas traditional shopping routines are quite well established and understood. Finally, the strategies of sellers can be very sophisticated, complex, and multi-dimensional. The upshot is that electronic commerce may not have straightforward impacts. Even so, we have illustrated how standard economic theory helps us understand some of the possible changes in shopping behavior as consumers go online.

Spending and Saving We now shift gears to look at time from another angle, that of saving for the future. We save for a rainy day, but also because we know that almost surely we will need some savings when we are no longer able to work. We focus on the latter motive, leaving aside the precautionary motive, since that requires a more complicated analysis. We want to continue to show how economic analysis can be used to describe individual life choices. We also want to lay the groundwork for considering the interesting question of what we do with our savings. Many changes in the world around us (particularly financial sector deregulation and concerns about social security) have made our savings decisions more important for us, while simultaneously, the Internet has opened up many new possibilities for implementing our financial decisions.

Suppose that Ernesto's life can be divided into two parts: his working life and a retirement period. He obviously needs to consume in both phases of his life, but he will earn only in the first phase. Suppose there is no social security program where he lives, so he has to rely on his savings. How much will Ernesto save? The key determinants of Ernesto's decision are his income, how he values consumption in the two phases of his life, and the rate of return he gets on his savings. Let us suppose that Ernesto has no discretion in how much to work, and that his income from working will be B dollars. This keeps the number of his choice variables at two, and lets us use the same kind of diagram as before. Also, suppose that the return he gets on his savings is 100r per cent. Thus, if r = 0.10, his rate of return is 10 %. We will denote Ernesto's expenditure on consumption in the two phases of his life by $C_{working}$ and $C_{retired}$, and both these are measured in dollars. Ernesto's budget constraint is given by:

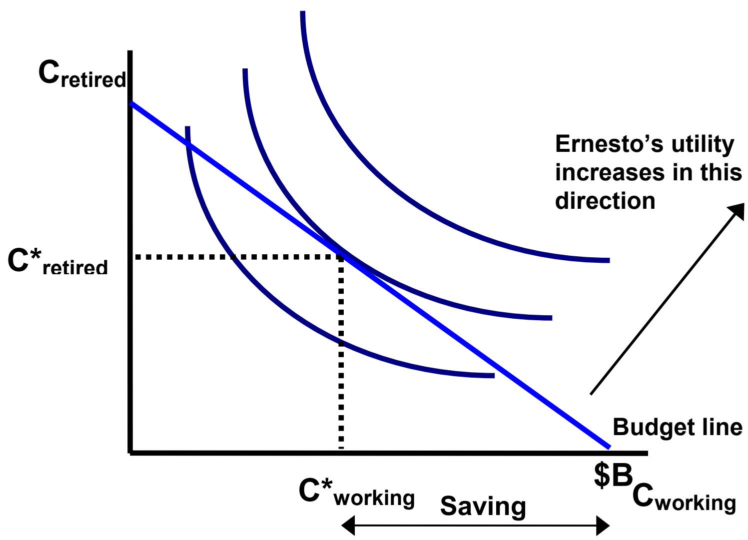
$C_{retired} = (1 + r)(B - C_{working})$.

The right hand side of this equation is how much he has available to spend during his retirement phase. It multiplies two things: his gross return per dollar saved, (1 + r), and his dollar savings ($B - C_{working}$). If we rearrange Ernesto's budget constraint, we get something that looks very much like Shawn's budget constraint as illustrated in Figure 4.14. The rearranged equation is:

$C_{retired} + (1 + r)C_{working} = (1 + r)B$.

Ernesto's situation is illustrated in Figure 4.18.

Figure 4.18



Putting aside savings may sometimes be difficult, but that has more to do with the peculiarities of our psychology. We may have difficulty deferring our satisfaction, and we may focus more on the near future than on a distant retirement. We often overcome these problems by committing ourselves to regular savings through automatic payroll deductions. When we do so, however, a more complex choice confronts us. What shall we do with our savings?

A bewildering variety of options awaits our savings dollars. We can buy bonds or stocks, or both. There are thousands of stocks to choose from. If we rely on mutual funds, which are bundles of stocks picked by professionals, we still have hundreds of mutual funds to choose from. While the practical process of deciding what financial assets to buy with our savings will involve a great deal of information gathering and assessment, the tools of economic as reviewed in this chapter provide us with an understanding of the most fundamental issue underlying choices of financial assets. In addition to providing us with a return on our savings, financial assets typically also involve risk. Also, assets with higher average returns involve more risk. As long as we do not enjoy bearing more risk (we ignored this in looking at Ernesto's savings decision, above), choosing how to allocate our savings involves optimizing the tradeoff between average return and risk. We illustrate this decision-making next.

Figure 4.9a illustrates three different assets, distinguished by their risk and average return combinations per dollar spent on them. These assets are the stocks of three companies, Highflyer.com, General Business, and Microprofit. Figure 4.9a, also shows Ernesto's indifference curves. These have the same shape as usual, but because Ernesto prefers less risk to more, given a particular average return, they point to the northwest instead of the northeast. In Figure 4.19b, we show Ernesto's utility-maximizing choice of how to allocate his savings. The straight line joining the risk-return combinations obtainable from Highflyer.com and General Business represents all different portfolios combining these two stocks. Since the risk-return combination obtainable from buying shares in Microprofit lies below this line, relative to the direction in which Ernesto's utility increases, he does not buy Microprofit stock.

Figure 4.19a

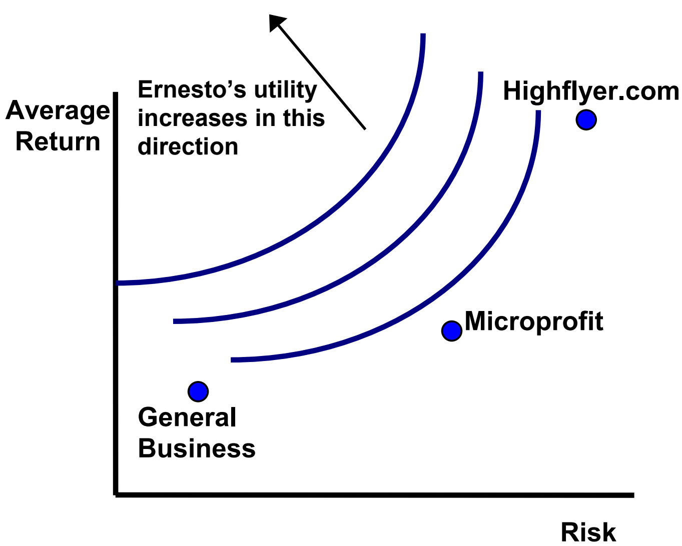
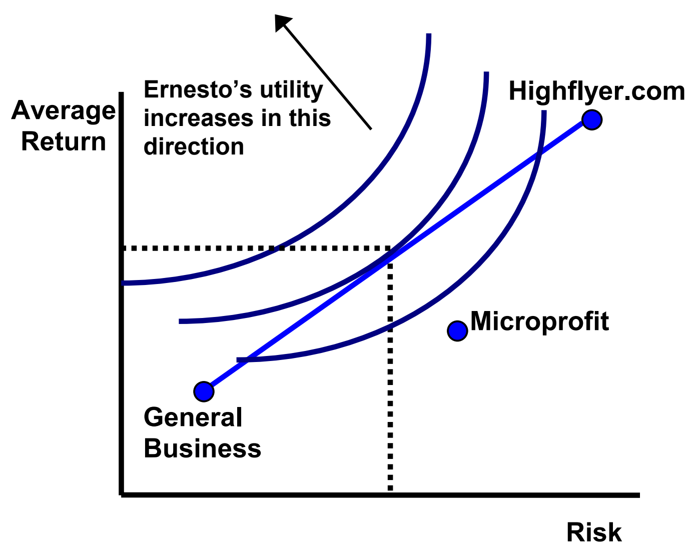


Figure 4.19b



Note that Microprofit's stock is not worse than either General Business's or Highflyer.com's individually. Its stock is riskier than General Business's, but provides a higher average return. Its stock gives a lower average return than Highflyer.com's, but it is less risky. Nevertheless, by diversifying between the stocks of General Business and Highflyer.com, Ernesto achieves a risk-return combination that is better for him than he could achieve with any portfolio that includes Microprofit.

We have described the fundamental economic insight in allocating savings among financial assets: achieving optimal combinations of risk and average return. Of course the problems facing the real saver are much more complex, involving thousands of possible assets, multiple time periods and decision-making points, and a constant flow of information about the assets, and events that will affect their characteristics. This information has to be gathered, processed into usable form, and analyzed in order to make sensible decisions. Formerly, only financial services professionals had access to this information. The limited information available to individuals, plus the limited savings options directly accessible to them, meant that we rarely confronted the details of decisions such as those illustrated in simplified form in Figure 4.19.

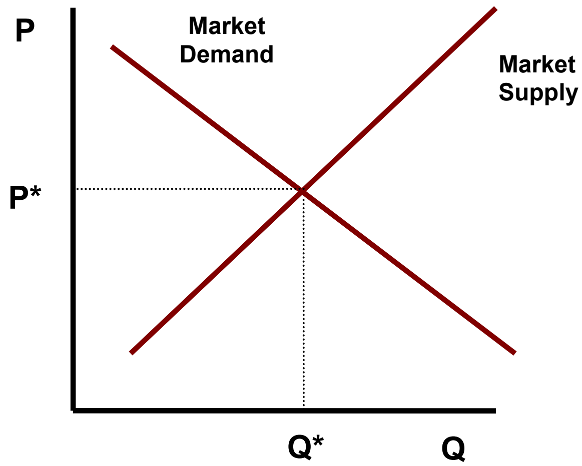
Financial deregulation, a slow and ongoing process that began in the 1980s in the United States as well as other countries, has opened up numerous options for individual decision-making about allocating financial assets. The Internet has turned out to provide the perfect tool to act on those options. Financial information is available in quantities unimaginable a few years ago, and now decision-making tools for processing and analyzing this raw information are also becoming available. Finally, since ownership of financial assets is merely an electronic record, financial actions can be taken purely online. Money, like music and news is a quintessential digital product. We will elaborate on the online financial service revolution in Chapter 20.

4.4 What Markets Do

Markets bring buyers and sellers together. A market can be thought of as a specific institution for achieving this goal, or simply as an abstraction, describing the interaction of buyers and sellers of a particular product. Ultimately, however, real institutions underlie the abstractions, and the rise of the Internet has focused attention on the rules of action and communication that constitute market institutions. We can think of many specific examples of markets, with their own rules. The Santa Cruz Farmer's Market takes place every Wednesday afternoon, in a particular downtown parking lot. Prices for produce are posted, and typically fixed. In a flea market, or a produce market in a developing country, on the other hand, you might expect to haggle over the price. A modern financial market may take place with geographically dispersed traders entering their 'bids' and 'asks' on desktop computers, according to prearranged rules for these offers. This illustrates the range of institutions that are encompassed in the term 'market'.

Let us begin, though, with the textbook abstractions. In a market with many sellers and many buyers, no single market actor alone has the power to 'move the market', to affect the price at which the product or service is bought and sold. The willingness and ability of sellers to provide quantities to the market is summarized graphically in a market supply curve, and the willingness and ability of buyers to purchase those quantities is summarized in a market demand curve. These are shown in Figure 4.20, in what is probably the most familiar and useful construct in all of economic thinking.

Figure 4.20



The intersection of the supply curve and the demand curve indicates the market equilibrium: the quantity that will be sold in the market, and the price at which it will be sold. At any other price, the argument goes, there will be pressure for the price to change. At a price higher than P*, there would be an excess supply (the quantity that would be supplied exceeds the quantity that would be demanded), and therefore the higher price could not be sustained. At a lower price than P*, there would be an excess demand, and upward pressure on the price as a result.

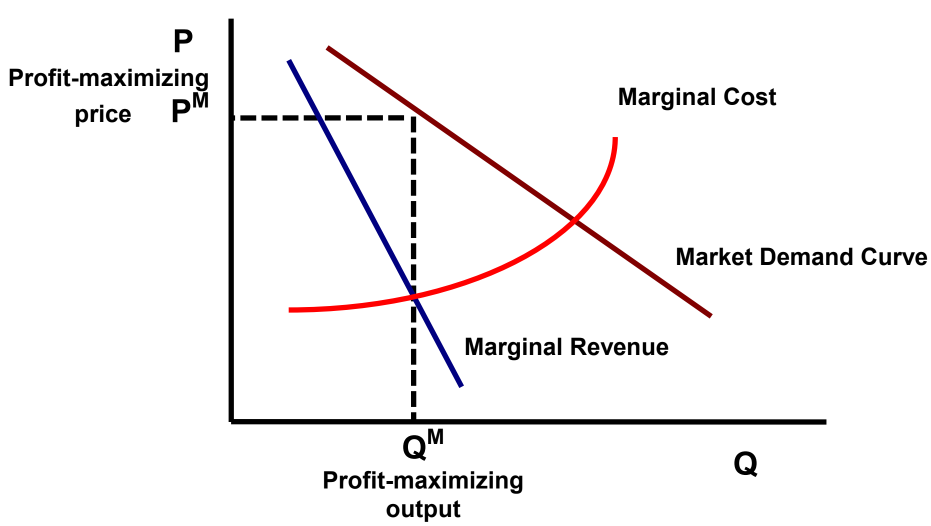
This story is a good first approximation, because it is the basis for a host of specific predictions about the working of markets. If the market in Figure 4.20 is the US market for oranges, we can use simple supply-demand analysis to analyze the overall effects of a freeze in Florida (the supply curve shifts to the left, price rises, quantity falls), a successful advertising campaign featuring celebrities drinking orange juice (the demand curve shifts to the right, price and quantity both rise), and other such events. In making these predictions, we do not worry about the details of how the adjustment of price and quantity actually takes place.

Why should it matter, as long as the final prediction is right? There are several possible answers, depending on the particular market. In some cases, the prediction of the textbook model may be qualitatively right, but get the magnitudes of changes wrong. This can be thought of as the result of 'frictions' in the process of adjustment to a new equilibrium. In other cases, these frictions may be sufficient to negate the predictions of the model. In particular, the frictions may create market power for sellers, requiring a different model altogether. In either case, understanding the source of market frictions is critical to understanding the potential impacts of electronic commerce in general, and electronic marketplaces in particular. Much of the potential of online commerce lies in removing frictions and making actual markets work more like the idealized textbook model.

This is a good place to indicate some of the directions we will take. Supply and demand curves are introduced with the idea that no single actor sets or influences the market price. This leaves the process of price determination somewhat of a mystery. In practice, of course, there is some method, typically involving human decision-makers, for determining prices. Buyers and sellers may haggle bilaterally, making offers and counter-offers. Sellers may post prices, and adjust them based on what they observe about actual quantities sold. There may be intermediaries that announce prices at which they will buy (bids) and at which they will sell (asks), which they adjust based on the responses they get from buyers and sellers. And so on. These processes are based on particular information that is observed by the price setters or market intermediaries, and the lack of perfect information is one source of friction in the market. You can guess where the Internet comes in to reducing the friction – it increases the amount of information available as never before.

It is useful to consider the market power aspect of market price determination more explicitly, since it also involves a standard economic analysis. Figure 4.21 reproduces Figure 4.7, showing the behavior of a monopolist. The market demand curve is the firm's demand curve, and it uses the MR = MC rule to determine its profit-maximizing quantity and price. These are then the market quantity and price. There is no supply curve in this case, because the firm does not have a strategy of offering to supply a particular amount at the going market price: it simultaneously determines the market price and the quantity it will supply. If the demand curve and marginal revenue curve are pivoted in a way that makes them steeper, but does not change the MR = MC point, then the monopoly firm's profit-maximizing quantity will be unchanged, but the price it charges will be higher. This could not happen in the case of competitive supply. Thus the predictions of the case where the seller has market power can differ from those in the competitive supply situation.

Figure 4.21



The point that a monopoly firm does not quite fit the supply-demand story also applies to any other situation with market power, that is, various types of oligopoly with strategic behavior against rivals. Within this class of oligopoly situations, also, there is variation in the prediction of what happens in the overall market. For example, if there are two firms and both set their quantities, as illustrated earlier in Figures 4.10 and 4.11, the market price is typically higher and the market quantity is lower than in the case where the same two firms compete by setting their prices (as illustrated in Figure 4.12). The intuition for this assertion comes from the logic of strategic substitutes and complements. In quantity-setting, the quantities are strategic substitutes, and a rival accommodates your aggressive behavior. In price-setting, the prices are strategic substitutes, and a rival will respond in kind to your aggressive behavior. As a result, price competition leads to a more competitive outcome (lower market price, higher market quantity) than quantity competition.

While oligopolies are more common than pure monopolies in the real world, the monopoly model has some interesting implications for understanding market institutions and outcomes. Bill Gates and his colleagues at Microsoft are emphatic in asserting that, even if Microsoft is a monopoly (which, to be on the safe side, they also deny), its monopoly is only temporary and fragile. At any moment, an innovative competitor can come along and take away chunks of Microsoft's market share. How might a monopolist faced with potential competition behave? It could wait till there is actual competition, and then cut its price in response. Or, under some circumstances (these are complicated to describe precisely), it might price lower than the monopoly profit-maximizing price, in order to deter entry. In the latter case, the possibility of competition in the future places some limit on the exercise of monopoly market power.

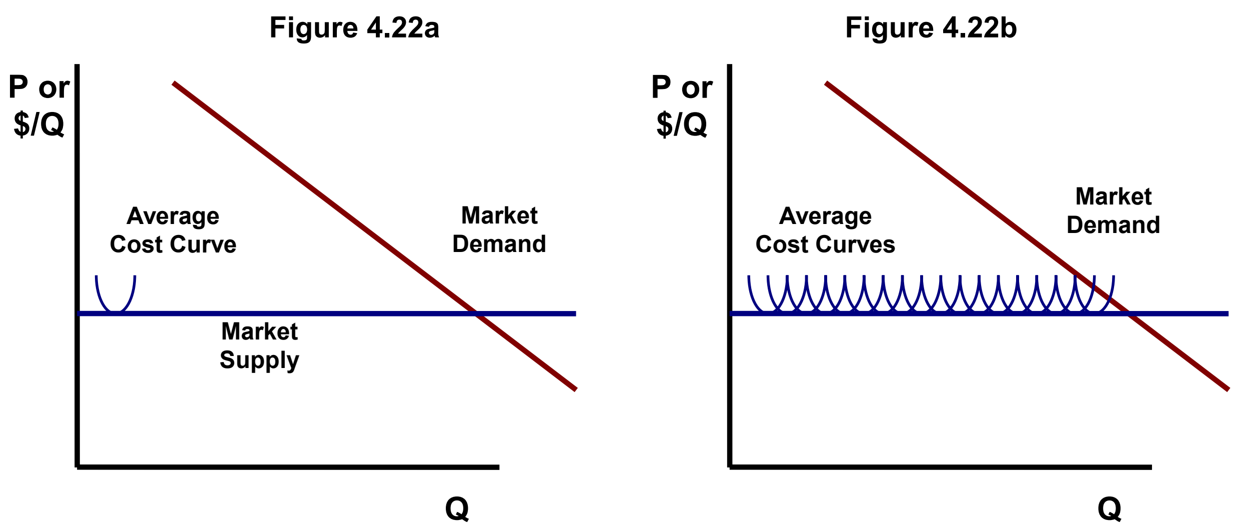
The idea of temporary monopoly power has some useful implications for understanding market institutions. Essentially, market frictions create (truly) temporary monopolies. As an example, consider a market where a buyer has to physically go sequentially from seller to seller. Each stage of this journey is costly, in terms of travel costs, time, or both. When a buyer is in a store, the seller has a temporary monopoly. The buyer knows there are competitors out there, in the near future (the time it takes to go to the next store), who might have lower prices, and the seller's knowledge of this possibility limits the monopoly power, but the situation is different than one where the buyer can simultaneously choose among competing sellers. Temporary monopoly resulting from frictions in market institutions therefore may lead to a different market outcome than the supply-demand prediction of Figure 4.20. If online commerce overcomes these frictions the market outcome may approach the competitive idealization. Note that the situation discussed here is a partial explication, from the seller's perspective, of the ideas explored in the discussion of shopping, where time spent searching could lead to a lower price paid.

.

4.5 Value, Scarcity and Economic Rent

Much of economics is ultimately about value, how firms create value for consumers, how this value is reflected in market prices, and how the capture of value is determined by market conditions. The economic approach to understanding the creation and capture of value is a major theme of this book, for the heart of the e-commerce revolution is in the changes it forces on the creation and capture of value. As one might expect, relative scarcity is an important factor, and we shall see in this section precisely how it matters. We shall also review the idea of economic rent – which represents excess value associated with a particular use of resources – and explain how it is driven by scarcity.

The basic supply-demand model provides a versatile framework for the discussion of these concepts. Figure 4.22 illustrates a particular market situation. In Figure 4.22a, the market supply curve is horizontal, meaning that any quantity will be supplied to the market at the price represented by the height of the supply curve. This could be because the marginal cost of supplying more is always at this level. This would be the result of constant returns to scale in production (output scales proportionately to inputs). The situation in Figure 4.22a could also be the result of the existence of many firms that are small relative to the market. Each firm has a U-shaped average cost curve, reflecting first increasing, then decreasing returns, but the overall effect of all these firms is the same as if there were constant returns.



In the market illustrated in Figure 4.22, many identical firms make up the industry supply. If the demand curve shifts to the right for some reason, then more firms will enter, with the same capital, skills and knowledge as the existing firms. If the demand curve shifts to the left, some firms will exit to maintain the market price. This market price is determined, therefore, by entry and exit, and by the minimum level of the average cost curves., which depends on the technology. By assumption, this technology is widely known, and not protected by patents, so any firm has access to it. Each firm in the industry earns zero economic profit, though they earn accounting profits that reflect the return on the capital used in production.

Is value created in this industry? The resources used in production here could be employed elsewhere just as effectively. Our assumptions mean that the capital used here could earn an equal rate of return in another industry. However, buyers are certainly better off as a result of the output they purchase. For every unit except the very last one, the buyer's willingness to pay exceeds the market price. This is illustrated by the demand curve being above the supply curve. The difference between the willingness to pay and the market price measures value that is created and that is captured by buyers.

Note that scarcity might change without altering the result that all firms in the industry earn zero economic profits. If some input becomes more expensive, all the average cost curves will shift up, and the market supply curve will also shift up. The product's scarcity increases, and this results in a higher market price, but not in a change in the situation of firms in the industry.

Now consider a situation where the conditions of supply are very different. Suppose that sellers have stocks of some product that can not be produced any more. For concreteness, think of some old toy or baseball card that is a collectors' item. Suppose that the sellers themselves do not put any value on the item, but the buyers, who are collectors of this category of item, do. If the item is worthless to sellers, they should be willing to part with it for any price that is above zero. However, we see in Figure 4.23a that they receive much more than that, thanks to the fact that buyers are willing to pay to add to their collections. For whatever psychological reasons, this gives them satisfaction. Buyers may also hope to resell it for a higher price later, but that leads to a more complicated story, because the question then is why sellers are willing to sell now rather than wait for the higher price.

Figure 4.23a

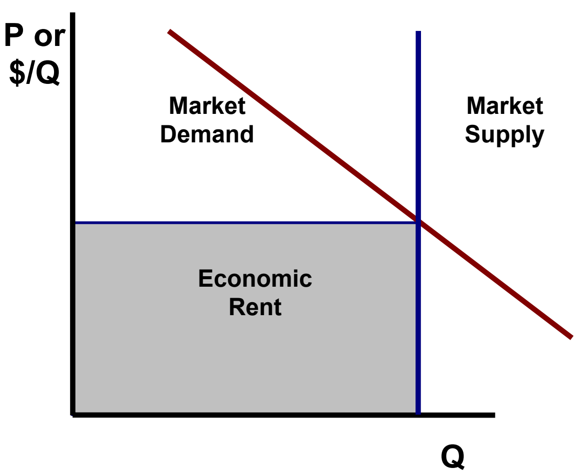
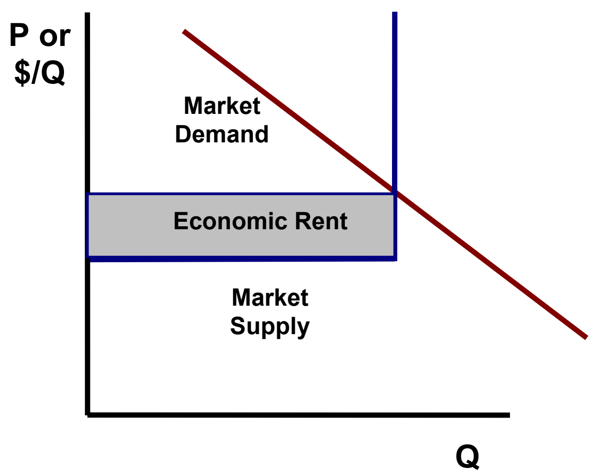


Figure 4.23b



The collectors items are worth much more when sold to the buyers than if kept by the sellers. The difference between the market price and the value of the items to the sellers is what we call **economic rent**. The total economic rent in the market is shown as the shaded rectangle in Figure 4.23a, and is here exactly equal to the sellers' total revenue. This value is captured by the sellers. The term 'rent' comes from the analogy with land, which, in David Ricardo's introduction of the concept in the 19th century, was in fixed supply and had no alternative use, just like the collectors' items here. Unlike the situation in Figure 4.22, therefore, sellers do better than with an alternative use of their resources (here just the product itself). The buyers are in the same situation as before: they get some surplus value, because their willingness to pay, which reflects the value to them of having the collectors' item, exceeds the market price.

We may briefly consider the effects of changing scarcity in this case. Suppose the supply curve shifts to the left in Figure 4.23a., so that the collectors' item is relatively more scarce. The economic rent rectangle is now taller but narrower. Whether the area increases or decreases depends precisely on how sensitive is the quantity demanded to changes in price. If the demand is price elastic, then the total economic rent will actually go down. Of course, the economic rent per unit sold must go up, and in that sense scarcity and economic rent are aligned.

Figure 4.23b shows a situation where the collectors' item is a new one. It is going to be produced, under constant returns to scale, but in a 'limited edition'. The constant returns to scale production implies a horizontal supply curve, as in Figure 4.22. The limit is shown by the fact that the supply curve becomes vertical at that point: no more may be produced. Think of many small firms are making a toy for Disney, which pays them their average and marginal cost. These firms, or any others, can not produce more than their orders from Disney, because otherwise they will get sued for copyright infringement, and they will lose future orders.

Disney's profit here is an economic profit, which is also an economic rent. In Figure 4.23b, the top of the economic rent rectangle is the market price, as in Figure 4.23a, while the bottom of the rectangle is now the payment to Disney's contract suppliers, which is the average and marginal cost of production. Notice that not limiting production will drive Disney's economic rent to zero, since then the rectangle will disappear. The optimal limit for Disney is one which makes the rectangle's area as large as possible. Where this limit lies will depend on the price elasticity of demand, that is, the responsiveness of quantity demanded to changes in price.

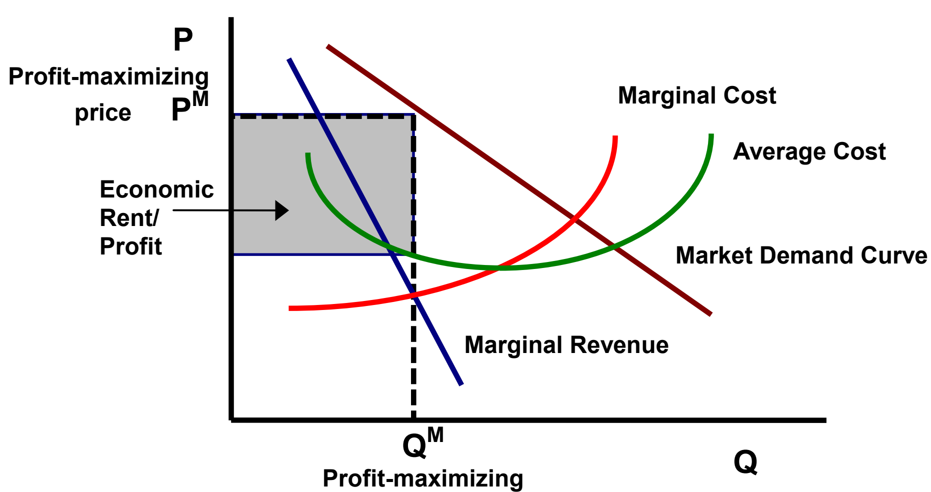
Disney's economic rent here can be thought of as a return to some unique input it controls. In this case, the unique input is the Disney brand name, and all the characters and other intellectual property that Disney owns. Since these can not be legally imitated, entry is not possible in the same way that it was in the situation illustrated in Figure 4.22. Disney's suppliers, on the other hand, control no such unique inputs, and they receive no economic rents at all. They are in the same position as the firms in Figure 4.22. The implication of this example for e-commerce (and any kind of commerce) is direct and obvious: economic rents will accrue only to those with control over unique inputs. Imitation and entry are the enemy of economic rents!

In the Disney example, economic rents are a return to a unique input in fixed supply: Disney's brand name and its stable of characters. These intellectual properties are protected by copyright law (the characters) and trademark law (the brand itself). In many cases, patent law provides the economic rents. Amazon.com right now seems a long way from earning even normal profits, let alone economic rents, but its stock market valuation must reflect expectations that it will do so in the future. Part of its strategy is based on building a brand name, part on locking in customers (see Chapter 16). However, despite Jeff Bezos' call for rethinking the law on software patents (see Chapter 3), Amazon has certainly not renounced its 'one'click' shopping patent. If such patents can be enforced, they act as a barrier to imitation and entry, and are at least a supporting element of a strategy that seeks as high profits as possible.

The fact that Disney or other content creators, as well as inventors of patentable innovations earn economic rents does not imply that these rewards are somehow unjustified. It simply involves a classification of rewards. Michael Jordan earned millions as arguably the greatest basketball player ever. His next best occupation appeared to be minor league baseball, with a much smaller salary. The difference between the two payments was a rather large economic rent, reflecting his enormous skill and his consequent popularity. One may hold different value judgments on the appropriateness of such large rewards, but they are quite separate from the classification of the rewards as economic rent. In Jordan's case, his performance was the result of native skill, hard work, and perhaps luck. These factors all enter into the innovation process that creates intellectual property. One difference is that, while no one could easily imitate or copy Michael Jordan's performance, intellectual property is easily subject to that kind of imitation or copying, and this is where the legal protections discussed in Chapter 3 have a role.

Temporary IP monopolies awarded by law to innovators and creators have a positive role in encouraging future innovation, in addition to the case that can be made on the grounds of justice. Other kinds of monopolies may be less desirable. In all cases, however, they involve the capture of value, and economic rent in particular. If we add the average cost curve to Figure 4.21, as we do in Figure 4.24, we can illustrate the monopoly firm's economic profit, which is an economic rent, to the extent that the same resources used elsewhere, in a competitive industry, would not earn as high a return. By assumption, entry into this industry is barred, and the economic rent is preserved. The monopoly still creates value for its customers, but the restriction of output and higher price mean that there is some loss to customers in doing so. The restriction of the production of the collectors' item in the Disney example involved a similar loss, though that might be counteracted if buyers get satisfaction from knowing that only a limited number of others have the item!

Figure 4.24



Note that protection of intellectual property need not lead to monopoly, or to a loss of buyer value. In Figure 4.25a, one of a number of small firms has an innovation that lowers its costs, and it earns economic rents as a result. The innovation does not lead to market dominance, because it is not associated with increasing returns to scale. The market price is not affected by the innovation. In Figure 4.25b, the same firm does better by licensing its innovation to all others in the industry. Since it has a monopoly on the innovation, it is able to charge license fees that extract all the cost savings from the licensee firms. The innovator earns economic rents, but the product market stays competitive, and the licensees all make zero economic profits. The assumption here is that license fees are extracted in a lump-sum fashion, say as an up-front payment. This means that the licensees' own output decisions are not distorted by any concern about license fees. Because the license fees are equal to the cost savings from the innovation, the licensees' average cost curves do not fall.

Finally, we return to the kind of monopoly, or market power in general, that is based on restricted entry. Temporary monopolies that result from the imperfections of market information possessed by buyers will also lead to economic rents. Another source of market power lies in middleman or intermediary roles, where entry is difficult. These middlemen, such as bond traders and other financial intermediaries, are like toll takers. E-commerce has the potential to erode economic rents in both these areas, by creating new highways that bypass these toll collectors. However, the economic rents will not necessarily accrue to new competing intermediaries, or to competing providers of market information. The lesson of our analysis is that only the owners of unique inputs, or the firms that are built around those inputs, are likely to earn economic rents in the new economy. Value creators do not always capture that value. The innovator in Figure 4.25 is precisely such a winning firm.

Figure 4.25a

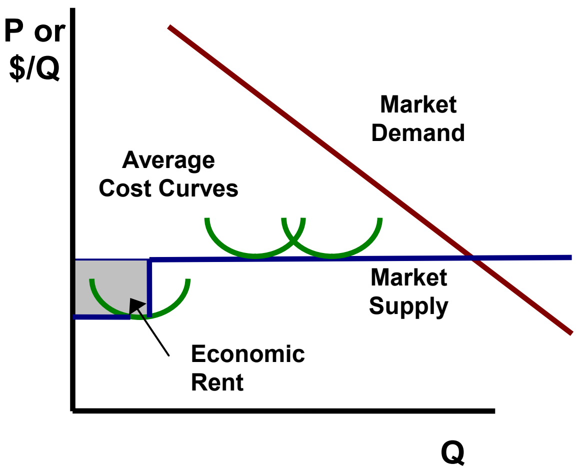
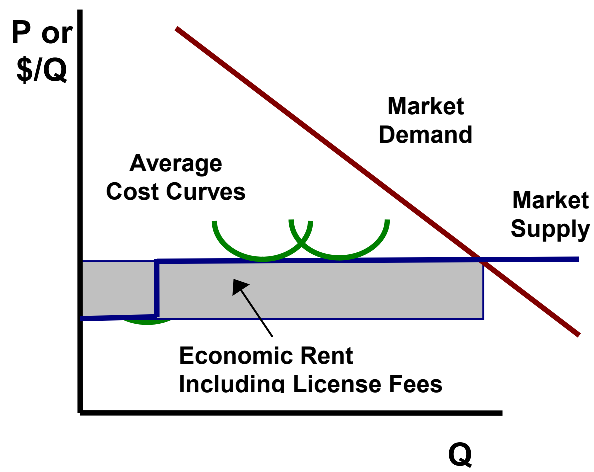


Figure 4.25b



4.6 Conclusion

Standard economic principles provide an excellent starting point for understanding the potential impact and development of electronic commerce. Profit-maximizing firms and utility-maximizing individuals, making their separate decisions and interacting in the marketplace, follow the same laws of economics that they always have. Production and exchange create value and share it between buyers and sellers, just as they have done for centuries. What is new? E-commerce focuses attention on the minutiae of the market – the processes of gathering market information, the rules that govern what is offered, the roles of those wh actually make markets function. E-commerce also highlights the roles of innovation and intellectual property, and how the rewards for innovation are decided in the marketplace, given a particular legal framework. Finally, e-commerce is about information, both as a product or service itself, and as an enabler of markets for other products and services. We have introduced some of these themes in this chapter, along with a review of a basic economic toolkit. The rest of the book awaits.

Summary

- Firms can be viewed in abstract as profit-maximizing entities that use technology to convert inputs into outputs. Key aspects of technology, particularly the idea of returns to scale, are captured in the shape of the average cost curve.
- Firms' market power depends partly on the level of output at which decreasing returns set in, relative to the level of market demand. Firms that are relatively small behave competitively, and free entry ensures that they will not make economic profits (returns higher than could be got in another industry) for long.
- Various types of strategic behavior are plausible, and market outcomes will depend on the precise nature of strategic behavior. A general concept for predicting behavior is the Nash Equilibrium, in which each strategic firm is choosing its profit-maximizing response to its rivals.
- The choices of utility-maximizing individual buyers can be described using indifference curves and budget lines: each individual makes a choice on the highest indifference curve that is affordable (lies on the budget line).
- The same approach can be extended to describe individuals' choices of leisure and working, how much time they devote to shopping and to searching for deals, how much to save, and how to allocate savings among different fiancial assets.
- Markets bring buyers and sellers together. Much of the innovation in e-commerce lies in changing the institutions that implement this process. Information flows

change, and the role of intermediaries in markets changes, often moving in the direction of more competition and lower frictions in the working of markets.

- Economic rents are returns above what can be earned elsewhere, and include economic profits as a special case. Economic rents may come from ownership of intellectual property or other unique inputs, or they may come from various other monopoly positions that involve restricted entry. E-commerce may erode the latter kinds of rents.

Questions

1. Do you think that e-commerce involves increasing returns to scale more significantly or less so than traditional bricks-and-mortar selling? Do high initial costs of building a web site replace the costs of building more and more physical stores? Or does e-commerce make it possible for small firms to set up shop and reach customers more cheaply? Will e-commerce lead to more competitive industry structures or less?
2. Is online shopping just a more efficient, more convenient kind of catalogue shopping? Why have traditional stores survived catalogue shopping? How have they responded to competition from catalogues? In what kinds of markets do retailers do well with catalogues? When do firms use both approaches?
3. Can being the first and biggest online retailer be enough to generate continued economic profits (that is, profits above normal rates of return)? What aspects of technology can support this outcome? What aspects of customer behavior can help?