5

How shall the firm produce the product?

Learning outcomes

If you study this chapter carefully, it may help you to understand:

- the law of diminishing marginal returns
- the concept of returns to scale
- what a production function is and what it is not
- different types of knowledge
- the importance of learning-by-doing
- the division of labour.

To the extent that you develop such understanding, you should be better able to:

- appreciate the difference between the mainstream theory of production and the heterodox approach
- identify the different kinds of knowledge needed in order to produce a good or service.

5.1 Introduction

In this chapter we assume that the entrepreneur is confident that consumers will want the product and we direct our attention to the issue of how to produce the good or service in question.

Production is defined as the process of using materials and factor services, in conjunction with technology, to create new goods and services.

The aim of production is to add value to the chosen inputs. The branch of economics that addresses how best to produce a good or service is called the **theory of production**, but before we discuss the technical details of this theory we will take a look at some of the practical issues that faced the British vacuum cleaner entrepreneur James Dyson in his attempts to get his revolutionary Dual Cyclone technology into production. We will then take a look at the mainstream approach to the theory of production before moving on to look at the modifications and extensions to this framework that have been constructed by heterodox economists.

It is probably fair to say that the mainstream story of production is somewhat easier to understand than the heterodox story by virtue of the fact that it makes several simplifying assumptions that, in turn, make it a good place to start developing your theoretical understanding of production. On the other hand, the heterodox approach tells a more sophisticated story in which many of the simplifying assumptions of the mainstream story are either relaxed or criticized and replaced. Ultimately, we hope that by presenting the Dyson story and then the theoretical stories in sequence you will develop a sound grasp of the issues a real world entrepreneur has to grapple with. One final point before we begin: this chapter is an essential prerequisite to the one that follows where we will be taking a look at the costs a firm incurs as a result of the answers it provides to the production question.

5.2 The story of Dyson vacuum cleaners

In these opening years of the twenty-first century it is probably fair to say that Dyson vacuum cleaners with their distinctive design, colouring and advanced Dual Cyclone technology are a familiar sight in all electrical retailers, department stores and mail order catalogues throughout the UK and, increasingly, the rest of the world.

Given the ubiquity of the marque, it is easy to forget that the 'Dyson phenomenon' is a relatively recent addition to the rather staid world of the vacuum cleaner market. The first model produced by entrepreneur James Dyson's own

UK factory was manufactured on 1 July 1993, but the story of Dyson's struggle to bring his technology to market begins way back in 1978 when he discovered why the dominant vacuuming technology was such a poor performer and arrived at a solution to the performance issue by making connections between domestic vacuum cleaning and the apparently unrelated technology found in a typical sawmill.

In 1978 James Dyson was already a successful designer-engineer-entrepreneur sitting on the board of a company called Kirk-Dyson, which produced and marketed the 'Ballbarrow' (an all terrain wheelbarrow) and other gardening products. One day, while he was vacuuming his house with an old reconditioned upright cleaner, he began to think about why it only seemed capable of pushing dirt and dust around the floor rather than sucking it up as it was supposed to do. He reasoned that it was probably because the dirt-collection bag was full so he replaced it, but performance was little better. Thinking that this was probably because the cleaner was so old he purchased a brand new cylinder cleaner. This worked much more efficiently, but only for a short while. Once again, reasoning that the dirt-collection bag must be full he replaced it and performance was restored. As before, the improvement was only transitory and soon the vacuum cleaner was as ineffective as the old reconditioned upright he had been using. Rather than replacing the bag again he emptied it and reused it. He was intrigued to find out that this did not improve performance, even briefly, in the same way that a brand new bag had done. After conducting a few more experiments Dyson concluded that vacuum cleaners with bags became inefficient very quickly because the pores in the bag were becoming clogged with dust and this was reducing its porosity to the air, which had a knock-on effect of reducing the power of the suck generated by the vacuum cleaner's motor-driven fan. At this juncture he did not have a solution to the problem of diminishing suction power, but then a similar problem involving reduced suction power became apparent at the Ballbarrow factory in its powder coating process and the seeds of an idea began to germinate.

The frame of the Ballbarrow was epoxy-coated for toughness. The first stage of epoxy coating was achieved via the rather messy process of spraying a mist of powder over the Ballbarrow frames with the excess powder (the stuff that missed the frames) being sucked onto a fine gauze mesh for collection and reuse. The suction power was provided by a fan housed behind the mesh, but as more frames were coated the build up of plastic dust on the mesh reduced the efficiency of the suction provided by the fan. This required production to be stopped at regular intervals throughout the day to clean the mesh and collect the excess powder for reuse.

Unhappy with having to stop production at such frequent intervals Dyson inquired of the mesh filter manufacturers if a more efficient technology was available. They informed him that they could supply a cyclonic filter, as supplied to sawmills, which extracted excess dust from the atmosphere without the need for a mesh barrier. Instead it used a large conical funnel (some 30 feet in height) to generate a vortex of high-speed air which spun dust particles to the edges where they lost speed and fell into a collection bag. This technology proved to be too expensive for Kirk-Dyson to purchase so James Dyson visited a nearby sawmill, made some sketches and built a cyclone filter for the Ballbarrow factory. Then he made the important **mental connection** between this relatively large-scale technology and the small-scale technology of the vacuum cleaner – a connection that nobody else had ever made.

Eager to turn the mental connection into a **practical connection** James Dyson returned home and produced a small prototype cardboard cyclonic filter which he attached to his old upright vacuum cleaner (after ripping off its bag). Then he vacuumed his house and discovered that this basic small-scale cyclonic filter worked very satisfactorily in its domestic application, just as he had hoped it would.

Having got the basic idea Dyson threw himself into designing and developing a perfect cyclonic filtration system. He was convinced he had a technology that would take the vacuum cleaner market by storm. However, to his surprise and dismay the road to success was going to prove to be far from smooth.

In early 1979 Dyson left Kirk-Dyson to pursue development of the cyclonic system full-time. In conjunction with an old business associate, Jeremy Fry, he set up the 'Air Power Vacuum Cleaner Company' and commenced the lonely process of technology development in a run-down coach house adjacent to his home in Bath, England. In late 1982, after three years of constant work, the single cyclone idea had been transformed into a more efficient double cyclone (patented as the 'Dual Cyclone'), which was capable of sucking up dirt particles of any size, from the microscopic through to the very large (not to mention items such as coins, etc.). During this period Dyson had constructed more than 1000 working models before he achieved 100 per cent efficiency of the system. The idea of the Air Power Vacuum Cleaner Company manufacturing a product based on the technology was mooted, but the financial position of the company prevented Dyson from following this path. Instead Dyson set up a new company called 'Prototypes Limited' and set out to license the technology to the big players in the UK vacuum cleaner market.

Amazingly, nobody was interested in the new technology except for a small company called Rotork with whom Dyson had been involved several years previously. In 1983 Rotork financed an operation whereby Dyson designed a vacuum cleaner, Zanussi manufactured it and Klene-eze sold it. All in all just 500 units (which they named 'Cyclon') were made and sold. This perhaps was a false start but it was a chance to **learn from experience** and, importantly, to obtain **customer feedback** with respect to function and durability. A picture of the Cyclon found its way into a TWA in-flight magazine, which stimulated international interest and there followed a protracted period of negotiations with a variety of vacuum cleaner manufacturers in the USA. By the middle of 1984 a licensing deal looked to be set with the American company Amway but, after handing over technical drawings and flying out to the USA to sign the final agreement, Amway decided they wanted to **renegotiate** the deal (this is something economists call 'hold up' and we will return to it later in the book). By the start of 1985 the Amway deal had collapsed and Prototypes Limited was getting very short on funds due to the costs of negotiation (mainly the fees of expensive lawyers to oversee the contract) borne to date. Furthermore, Amway had decided to sue for various reasons. Prototypes counter-sued and the legal battle ended in early 1985 when Dyson handed back the front money paid to him by Amway, who had to return his patents and terminate their licence agreement.

Fortunately for Dyson the Cyclon had also stimulated interest in the Dual Cyclone technology in Japan from a company called Apex Ltd. A deal was struck and James Dyson spent much of 1985–6 in Japan designing a vacuum cleaner called the 'G-Force' (which retained many features of the Cyclon) which he saw all the way through to production. Yet again another **learning experience** for Dyson and a much-needed financial lifeline. As he recalls: 'In the year I spent with the Japanese, I learnt an awful lot about design that would stand me in great stead when I set up alone to make the Dyson Dual Cyclone.'

In March 1986 the G-Force hit the Japanese market at a retail price of £1200. Despite this astronomical price it became a hit in Japan and within three years it was making sales of £12 million a year, although thanks to the terms of his licensing agreement Dyson only received £60 000 a year.

The success in Japan paved the way for a renewed assault on the North American market and in July 1986 a deal was reached with a Canadian company called Iona for a dry shampooing machine to be called 'Drytech' (ostensibly a Dual Cyclone vacuum cleaner, but in order to get around certain legal restrictions placed on Iona it had to be redesigned to deliver a powdered carpet shampoo). Back at his coach house, Dyson gathered a small team around him consisting of 'a couple of designers, recent graduates of the RCA, an engineer and a draftsman', in order to design the dry shampooer. This went on sale in June 1987.

In the background negotiations for a 'pure' vacuum cleaner to be called the 'Fantom' were in train, but just as the deal was about to be signed Iona informed Dyson that their buyer, Sears, had already got a dual cyclone vacuum cleaner in their stores which was manufactured by none other than Amway! Unsurprisingly, Amway's cyclonic vacuum cleaner infringed several of Dyson's patents and once again a legal battle followed with the help of Iona, although at a cost to Dyson of a renegotiated (i.e. worse) deal for the Fantom. This legal battle would rumble on for several years at an annual cost of £300 000.

By 1989 the Dual Cyclone technology was becoming well known. In 1990 Dyson signed a worldwide licensing deal with Johnson Wax to apply the technology to a tank vacuum cleaner and a back pack vacuum cleaner for sale to industry and commercial cleaners. Early this same year Vax got in touch with Dyson and asked him to build an upright vacuum cleaner for the UK market and, as Dyson put it: 'My little staff at the coach house and I were designing vacuum cleaners, and other bits and bobs, for our current licensees, and the Vax deal looked like the final piece of the jigsaw.'

It had always been Dyson's intention to produce an upright cleaner for the UK market so the Vax proposal was attractive. After signing the agreement and taking receipt of £75 000 front money Dyson and his team delivered the final design and drawings to Vax at the end of 1990. It soon became clear, however, that Vax were not entirely happy and they asked for a number of redesigns. By July 1991 it didn't look like they were going to commit to production so Dyson parted company with Vax. He was now keen to produce a vacuum cleaner for the UK market himself. The major problem he faced was obtaining finance. The Amway legal battle was still rolling on and it was diverting much-needed funds away from the company, but, just as he was despairing of reaching an agreement with Amway, the case was resolved and the financial burden removed. Dyson and his team launched into designing their first UK upright vacuum cleaner.

The product that was to become the DC01 took as its point of departure the G-Force. The design team worked in the now-refurbished coach house for over nine months. Computer-aided design was carried out upstairs while prototype models were built downstairs. According to Dyson: 'We made hundreds of little technical improvements to the cyclone, and to the cleaner head, and at the same time concentrated on reducing the number of screws, joins, and parts in the finished design – this would concentrate our minds on the essence of the function, and force the form to follow it most efficiently, and also, most importantly since *we never knew how much money we were going to have to put the thing into production, it could keep down the amount of tooling that we needed* (at about £20 000 a mould, any part we could possibly do without was rejected...).' [emphasis added]. On 2 May 1992, Dyson's 45th birthday, the team delivered 'the first, fully operational, visually perfect, Dyson Dual Cyclone'.

With a fully working demonstrator model in place Dyson now needed to obtain finance so that his newly formed company 'Dyson Appliances' could invest in the required production tooling required to mass produce the Dual Cyclone. Merchant banks proved to be sceptical towards the project and in the end he obtained a loan of £600 000 from the high street lender Lloyds Bank, secured on his London and Bath houses. He raised another £750 000 by selling all rights to Alco (who had taken over the Apex licence) for production of the G-Force in Japan.

Having sorted out the finance, Dyson travelled to Italy where he negotiated directly with 18 toolmakers, whom he had first dealt with via Zanussi when he was involved with the Cyclon project, for the production of 40 large moulding units (some of which weighed up to two tonnes). The tooling, which cost £900 000, was ready by the end of November and it arrived in the UK at the end of December 1992 where it was transported to a newly set up American-owned company called 'Phillips Plastics' who were based in Wrexham, Wales. Dyson had decided to contract out the moulding of the vacuum cleaner's parts and its assembly to Phillips rather than take on direct control of the production process. By the end of January 1993 the first Dyson DC01s rolled off the production line and fulfilled delivery contracts that Dyson had negotiated earlier in 1992 with the catalogue-based mail order shops Great Universal Stores (GUS) and Littlewoods. Orders had also been received from several chains of regional electricity board shops and John Lewis department stores and, in April 1993, a large order from the electrical retail chain Rumbelows was received. Things were looking good, but all was not quite right at Phillips Plastics. The first problem was a very poor system of **quality control**, so Dyson placed five of his own employees in the factory to **monitor** production practices to ensure that substandard cleaners did not find their way to the market. By the end of May about 12 000 vacuum cleaners had been sold when problem number two arrived in the form of a visit from Phillips's senior managers from the USA. They had decided that they wanted to renegotiate the original agreement with Dyson to allow them to double the assembly costs and increase the price of manufacturing the plastic parts (with Dyson's own machines, remember!) by 16 per cent. By this stage Dyson had had enough of such behaviour and he severed the relationship with Phillips after a hasty bout of legal action. What followed was an impressively quick reorganization that saw Dyson contracting out manufacture of the parts of the DC01 to various suppliers around the UK and Europe (which meant relocating his moulds and negotiating new contracts) and setting up his own assembly plant. This plant was located in an old Royal Mail depot in Chippenham, Wiltshire, England. The depot had floor space of 20 000 square feet.

On 1 July 1993 Dyson's efforts were rewarded when his first own-built DC01 rolled off his new Chippenham assembly line. Assembly of the DC01 was carried out entirely by hand and within a fortnight the new production line staff (14 in total) had output up to 100 DC01s a day with excellent quality control. The labour force was not the only thing that was up to speed; demand for the DC01 with its superior vacuuming technology soared so much that by February 1995 the Dyson had overtaken established companies such as Hoover and Electrolux to become the UK's top selling upright vacuum cleaner by volume. In addition, Dyson brought out his first cylindrical vacuum cleaner, the DC02, in 1995. This proved also to be a runaway success with the buying public and is illustrated in Figure 5.1.

In an attempt to keep pace with the rapid increase in demand Dyson expanded his labour force (there was no shortage of willing workers and he offered a good remuneration package) and made the most of the land at his disposal on the Chippenham site by assembling six Portakabins, four containers and a 10 000 square-foot tent to accommodate the goods inwards and higher levels of production called for by the explosion in demand. However, despite these ad hoc arrangements,

Figure 5.1 Earl cleans up with a Dyson
(Photograph by G. Rosales-Martinez)

Dyson hit a constraint in the land he could use and the Chippenham facility proved unable to produce more than 30 000 units per week. This was not enough; Dyson needed to find a new assembly plant.

In August 1995 production was moved to a 90 000 square-foot factory in Malmesbury, Wiltshire (just a few miles away from Chippenham). By 1996 the company's turnover had swollen from £3.5 million in 1993 to £85 million and Dyson Appliances was officially recognized as the fastest-growing manufacturing company in the UK. Again the success continued in the UK market and overseas and despite 12-hour shifts being the norm for the production line workers Dyson hit a constraint in production once more. After utilizing the 20 acres of land at his Malmesbury site as best he could (he opened a £20 million extension to the factory in 1998) Dyson attempted to expand facilities further by buying the field adjacent to the factory. Unfortunately the objections of local residents prevented this and in September 2002 Dyson relocated vacuum cleaner manufacture once again, but this time he left the UK and set up production facilities in Malaysia.

Lessons from the Dyson story

It is clear from this brief history of James Dyson's efforts that turning even the best of ideas into a final good or service requires the entrepreneur to display a dogged determination and a degree of **organizational ability**. It involves the need to **negotiate contracts** with suppliers, the need to **find a suitable production facility** and the need to **identify a suitable production technique**. It also requires the entrepreneur to **monitor** the **quality** of the inputs, the **effort levels** of the factor services used and of course the quality of the final good or service. None of these activities is costless and all of them take place in an environment that displays a fairly high degree of **uncertainty**. Furthermore, the **knowledge** of how to carry out each of these activities is not easily obtained, but it is susceptible to improvement over time as the entrepreneur goes through a **learning process**. In short, the Dyson case illustrates that the answer to the question posed in the title of this chapter is multifaceted and far from straightforward. How then does the economic theory of production approach the topic?

5.3 The mainstream theory of production

5.3.1 Overview

In order to introduce you to the mainstream theory of production we shall tell a simple story that emphasizes its key features. You will notice as our story unfolds that much of the detailed richness of the Dyson tale we have told above seems to be missing. This is not because neoclassical economists walk around with their eyes closed, nor is it because they do not appreciate the complexity of real world production issues! Instead it reflects the mandatory assumption that lies at the heart of all formal neoclassical models that all decision makers (entrepreneurs) are as equally knowledgeable as each other and as equally well informed about relevant issues. In addition, and as we have mentioned before, Ockham's razor has been applied vigorously to formal neoclassical models which has enabled neoclassical economists to develop an elegant body of formal theory which can be expressed concisely in the language of mathematics.

This does not mean that the neoclassical theory of production is incapable of providing some useful practical insights, but it has resulted in a theory of rather limited scope because it concerns itself exclusively with a **quantitative analysis** of the **conversion process** by which various amounts and combinations of inputs are converted into particular quantities of output with the aid of given quantities of labour and capital. In fact, as you will discover, the neoclassical theory of the firm assumes that the entrepreneur has already answered questions about how best to organize labour and capital and it focuses exclusively upon the quantitative input–output relationship that results. The input–output relationship is described by a conceptual tool called a **production function**.

The neoclassical theory of production, and by association the neoclassical theory of the firm, is often called a **black box theory**. This is because it provides no details about 'soft' or qualitative issues that are internal to the firm, such as how the firm is structured (e.g. its hierarchy) and the nature of its culture. In addition, it ignores the role of the entrepreneur and management and as a result it sheds no light upon the nature of knowledge in the firm. You will discover that issues related to the identification of suitable sources of supply, negotiation of contracts and quality control, all of which we have identified as important issues faced by Dyson, are assumed to have been satisfactorily dealt with by the entrepreneur, so they do not feature in the analysis either.

5.3.2 The basic elements of the mainstream story

Given what we have said about the limited scope of the neoclassical theory of production it is useful to begin by extracting the elements of the Dyson story that are reflected in the theory. As you might expect, these elements relate to the quantitative elements of the story so our attention will focus primarily upon the reasons given by neoclassical theory to explain why Dyson Appliances and other firms face restrictions on the quantity of output they are able to produce.

The part of the Dyson story that is most relevant from the perspective of neoclassical production theory occurs from 1993 onward, i.e. after Dyson had set up his own vacuum cleaner assembly plant. During this period the story illustrates that demand for Dual Cyclone vacuum cleaners expanded quite rapidly and in an attempt to ensure that production could keep pace with this exploding demand Dyson expanded production in **two distinct ways**.

The **first** way Dyson found to expand production was by employing increasing numbers of people (labour) who were organized into 12-hour shifts to work within the space available at the given sites. When this happens the firm is trying to make the most of its **fixed factor input** (land in our example, which of course means that factory size – that is capital – is necessarily restricted also). Economists characterize this way of increasing (and decreasing) the firm's output as taking place in a **short-run** decision period. This terminology can be quite confusing when you meet it for the first time because it is not defined in historical units of time (e.g. seconds, minutes, hours, days, etc.).

The short run is defined as a decision-making period during which the quantity of at least one of the factor inputs to the production process is fixed.

The short-run analysis of Dyson Appliances illustrates one of the most famous 'laws' of economics: the **law of diminishing returns**. For example, at the Chippenham site, James Dyson attempted to increase production by expanding the input of labour and making full use of his other readily available inputs, but eventually he faced a constraint upon further expansion brought about by a physical shortage of land. This meant that even if he had added more people to his labour force he simply would not have been able to build a factory large enough to accommodate them and they would therefore have been able to contribute very little to the firm's output. We will illustrate and define the law of diminishing returns in more detail below.

The **second** way Dyson used to expand output was by changing the quantity used of **all of the factor inputs** into the production of vacuum cleaners, e.g. a bigger plot of land at Malmesbury was obtained, which allowed a bigger factory to be built to accommodate an increased number of workers. If the entrepreneur/management team wishes to avoid diminishing returns but has no access to productivity-enhancing innovation or learning effects, it has no choice but to expand all of its factor inputs, in particular the factor input that is giving rise to the output constraint. This way of expanding output takes place in something that economists call a **long-run** decision period.

The long run is defined as a decision-making period during which the quantity of *all* of the inputs to the production process can be varied.

The distinction between the long run and the short run is important and it leads to two complementary theories of production: the **short-run theory of production** and the **long-run theory of production**.

The short-run theory of production examines the implications for the quantity of output of combining different quantities of a variable factor input with a given quantity of fixed factor input(s). The long-run theory of production examines the implications for the quantity of output of combining factor inputs when all of them are variable in quantity; in this situation the firm does not face a constraint on output caused by one of the input factors as it does in the short run, so the long-run problem for the entrepreneur is to decide how large the firm's output capacity should be. As you might imagine, it is useful to have some expectation of the likely size of the market for your product when making the long-run decision.

5.3.3 The short-run theory of production

The short-run theory of production revolves around the concept of diminishing returns that we introduced briefly above. It also requires us to assume that production of an output is best carried out by a team of factor inputs rather than by individuals striving to produce a particular output entirely on their own (more on which we will discuss a little later). To illustrate the theory we will use a simple example based upon the decisions a fictional entrepreneur makes when faced with a particular short-run production function.

Art and craft teacher Samantha Pinewood is looking for a new challenge. Despite earning a weekly wage of £150 as a teacher and pleas by the principal of her school to stay on Samantha has decided to change career and try her hand at running a business that manufactures ready-to-assemble coffee tables. With an offer from her former principal of re-employment if the business does not work out Samantha sets up 'Tables 4U', a small firm that will manufacture basic unfinished wooden coffee tables in boxed flat-pack form. Each flat-pack contains a simple (unstained, unvarnished) wooden table top, four wooden legs and a small packet of screws and glue that can be used by the final customer to assemble the table.

Tables 4U's production facilities consist of a small workshop (300 square feet) with enough space to accommodate three workbenches (a lathe, a bandsaw and a packing machine). We say it has **three units of capital (K)**, and we will assume that it cannot change this quantity of capital in the short run. Tables 4U has no problem getting hold of its raw materials from its suppliers who are happy to provide as much wood as it can use. The packets of screws and glue are bought in from an outside supplier also. Once again this supplier is able to provide as many packets as Tables 4U requires. Tables 4U can **easily hire as many units of labour (L)** as it requires at any time. The relationship between the relative quantities of factor inputs (K and L) and the weekly output of coffee tables is described by the **short-run production function** shown in Table 5.1.

A production function is a way of presenting the quantitative relationship between factor inputs and the *maximum output* attainable given the current state of technological knowledge.

In other words, it shows a list of '**ingredients**' (factors of production) required by the production process and it tells us in what quantities these factors can be combined to produce a maximum output by an entrepreneur who understands the latest techniques of production, i.e. **the state-of-the-art**. The assumption that the entrepreneur is *au fait* with the latest technological understanding might seem like a rather strong assumption to make, but it is of course nothing more than a natural extension of the assumptions made about human decision makers in all neoclassical models. We will return to this point in Section 5.4 below.

Samantha has just won Tables 4U's first contract. This is to supply Pinewørld, a UK-based chain of Swedish furniture stores, with 450 flat-packs a week. The question Samantha faces is, given that capital is fixed in quantity at three units, how many people should she employ to ensure that she can fulfil Pinewørld's order? Inspection of Table 5.1 reveals that Samantha will need to employ three units of labour.

If you examine the information contained in Table 5.1 you will notice that each successive unit of labour added leads to a **change in output** that is **increasing** until we add a fourth unit. For example, when the second person is employed

Table 5.1 **The short-run production function for Tables 4U when K=3 units**

Quantity of capital (K)	Quantity of labour (L)	Number of flat-packs per week (output, or total product, TP)
3	1	100
3	2	250
3	3	450
3	4	550
3	5	600
3	6	480

output increases from 100 flat-packs per week to 250 flat-packs per week, giving an increase in output *caused by* the introduction of the second person of 250 – 100 = 150 flat-packs. The change in output *caused by* adding the third person to the labour force is 450 – 250 = 200 flat-packs. The name we give to the change in the number of flat-packs that we attribute to each successive unit of labour employed is the **marginal product of labour (MP_L)**. So, the MP_L of the second person employed by Samantha Pinewood is 150 flat-packs.

The marginal product of any input (e.g. labour) is the *addition* to total output that results from employing *one extra unit* of the variable factor input.

If MP_L is increasing then we say that there are **increasing marginal returns** to the variable factor input. That is, every time we make a marginal change in the factor input (i.e. a one unit increase) our total output increases by more than it increased when we added the *previous unit* of the variable factor input. How can we explain this phenomenon? Three potential explanations of increasing marginal returns are possible.

The first explanation is that the second worker is more skilled at using the workbenches than the first worker and the third worker is more skilled than the second worker. The second explanation is that each extra worker added works harder (i.e. puts in more effort) than the previous worker. Both of these explanations are plausible; for example, you probably have friends and colleagues whom you consider to be more skilful and/or harder working relative to yourself when it comes to carrying out various jobs. However, the mainstream theory of production does not resort to either of these reasons to explain increasing marginal returns. In fact the mainstream theory assumes that each unit of labour works equally as hard as the other units of labour and is as equally skilled. The correct terminology here is to say that **labour** is considered to be a **homogeneous** input with respect to its level of effort and skill.

The third explanation of increasing marginal returns revolves around the concept of **team production**. Having a team of labour inputs allows the entrepreneur to **organize** the production process in such a way that each team member does not carry out every single task required to produce the good or service. Instead, the entrepreneur can divide the total number of tasks between the members of the labour force and in this way gain productivity benefits. The benefits arise because each member of the labour force is able to **specialize** in a subset of tasks rather than having to execute all of the tasks associated with the production of the final output. Economists call this concept **the division of labour** and we will explore it in greater detail below.

Increasing marginal returns do not last for long with the production function we have shown here which implies that there are limits to the practice of dividing labour in the short run. We can see that this must be true because Table 5.1 tells us that the marginal product of the fourth unit of labour is only 100 flat-packs compared to the MP_L for the third unit at 200 flat-packs. The MP_L of the fifth unit of labour is just 50 flat-packs. We can say, therefore, that at levels of employment beyond three units of labour **diminishing marginal returns** are present in our short-run production function. The phenomenon of diminishing marginal returns is believed to be present in all short-run production functions so economists have granted it the status of a 'law'.

The *Law of Diminishing Marginal Returns* states that, if successive units of a variable factor input are combined with a given amount of fixed factor inputs then *beyond some point* the addition to output will begin to decline.

The full results of increasing labour by one unit at a time are shown in Table 5.2 where we have recorded the marginal product of labour, MP_L, and average product of labour, AP_L (which is found by dividing the firm's output by the total number of workers employed). Note that we have recorded the marginal product of labour in the table in between the units of labour input and output. This is because **marginal product** is a measure of the **rate of change** of output, that is, it should be recorded at the point of transition between the previous level of output and the next level of output.

Table 5.2 **Short-run production function showing increasing marginal returns initially followed by diminishing marginal returns**

Quantity of capital (K)	Quantity of labour (L)	Number of flat-packs per week (TP)	Marginal product of labour (MP_L)	Average product of labour (AP_L)
			100	
3	1	100		100
			150	
3	2	250		125
			200	
3	3	450		150
			100	
3	4	550		137.5
			50	
3	5	600		120
			–120	
3	6	480		80

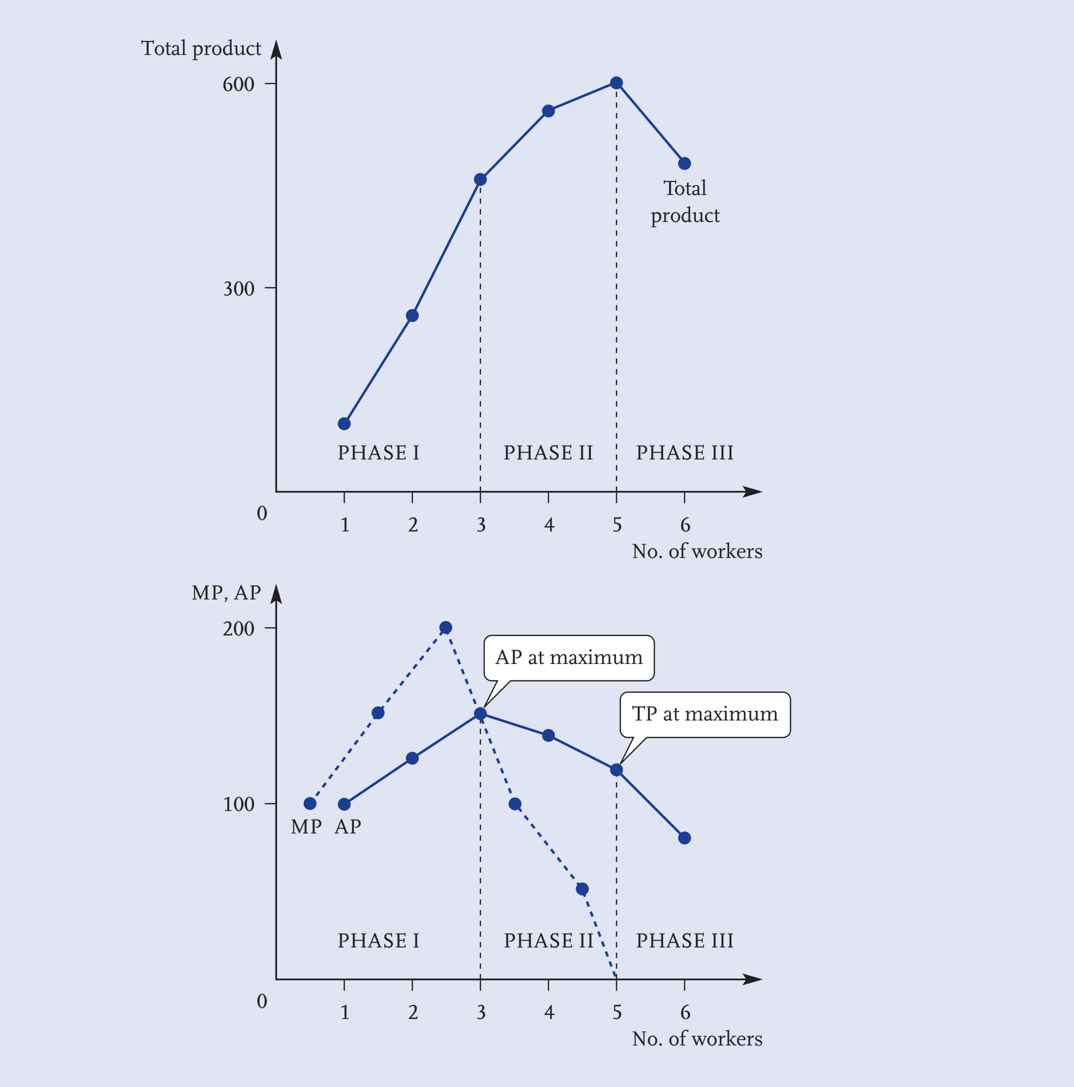


Figure 5.2 Total, average and marginal product at Tables 4U with quantity of capital fixed at three units

As you can see, diminishing marginal returns take a particularly strong hold if we add a sixth worker; in this case total output falls and we discover the **capacity limit** of the flat-pack workshop. This is useful information for Samantha to have, because if her business grows she will know the maximum output she can produce with her current quantity of capital. The MP_L, AP_L and total product (TP) are represented graphically in Figure 5.2.

Of particular interest is the relationship between MP_L and AP_L. You will observe that the plot of AP_L is at its maximum point where it intersects the plot of MP_L. We can use this to help us identify **three phases of production** where each phase represents how effectively the firm is using its fixed factor inputs:

- **Phase I**, which has an upper bound at the intersection of AP_L and MP_L, represents an **underutilization of the fixed factor inputs** in the production process, because increasing total product (output) within this phase leads to an increase in the AP_L and this can be interpreted as an increase in the effectiveness with which the firm is using its resources.
- **Phase II**, which begins at the boundary with phase I and ends where $MP_L = 0$, represents an increase in total product (output) as more labour is employed although now the firm is using its resources less effectively (as measured by the declining AP_L). We can say that compared to phase I, a firm that finds itself in phase II is **using its fixed factor inputs 'properly'** (i.e. it is not underutilizing them).
- The firm will not wish to find itself in **phase III** because here it has gone **beyond the capacity of its fixed inputs**. We can say, therefore, that a rational firm will want to operate somewhere in phase II, although we cannot say precisely where.

If the demand for the firm's product grows and takes the firm to the boundary between phase II and phase III the entrepreneur will need to expand the firm's productive capacity. This requires an understanding of the long-run theory of production.

5.3.4 The long-run theory of production

The long-run theory of production makes use of **two conceptual tools**. The first tool is something called an **isoquant map**. The second tool, which is used in conjunction with the isoquant map, is called an **isocost map**.

You will recall that a short-run production function tells us how output will change if we add successive units of a variable factor input, which in our example was labour, L, to a fixed factor input, which in our example was quantity of capital, K. In principle we could produce a short-run production function for every conceivable value of the fixed factor input and if we assume, for the sake of simplicity, that there are just two inputs, we can present this collection of short-run production functions in tabular format to obtain a **long-run production function**.

A long-run production function tells us the quantitative relationship between output and factor inputs when the quantities of all of the factor inputs in the production process are variable.

If you look at Table 5.3 you will see that we have produced an example of a long-run production function.

The two axes of Table 5.3 record the quantity of inputs used to produce the output figures contained in the body of the table. For example, the table tells us that 6 units of factor input Y can be combined with 4 units of factor input X to produce 673 units of output. If we fix the quantity of Y at 6 units we can read across this row in the table to find out how output varies in the short run as we add successive units of input X. To maintain consistency with the story of the short-run theory of production we have told in the previous section we will call the factor inputs X and Y labour, L, and capital, K, respectively, but you should note that this is not an essential element of the analysis because the theory of production can be applied to any two things that combine together to produce a good or service. For instance, input Y could be quantity of farmland (measured in hectares) while input X could be quantity of dairy cows and the relevant output quantity of milk (measured in litres). Alternatively, Y could represent the number of telephones in a call centre, X the quantity of telephone operatives (those annoying folk who call you up and try to sell you double glazing/insurance/kitchens, etc. at the most inconvenient of moments) with the relevant output being the total number of calls possible.

If you look carefully at Table 5.3 you will see that some of the output figures recorded in one cell are repeated in other cells. Take 230 units of output as a case

Table 5.3 An example of a long-run production function

Quantity of input Y (e.g. Capital, K)										
10	398	603	770	915	1046	1167	1280	1386	1487	1585
9	374	566	722	859	982	1095	1201	1301	1397	1487
8	348	528	673	800	915	1020	1119	1213	1301	1386
7	321	487	621	738	844	942	1033	1119	1201	1280
6	293	444	566	673	770	859	942	1020	1095	1167
5	263	398	508	603	690	770	844	915	982	1046
4	230	348	444	528	603	673	738	800	859	915
3	193	293	374	444	508	566	621	673	722	770
2	152	230	293	348	398	444	487	528	566	603
1	100	152	193	230	263	293	321	348	374	398
	1	2	3	4	5	6	7	8	9	10

Quantity of input X
(e.g. Labour, L)

in point. The production function tells us that **230 units of output** can be produced by combining the inputs in **three alternative ways**. The alternative input pairs that will give us 230 units of output are: (4Y and 1X), (2Y and 2X) and (1Y and 4X). We have highlighted in blue the cells containing 230. We have also highlighted in blue the cells containing 528 units of output. We are now in a position to introduce you to an isoquant map.

An isoquant is a line plotted on a graph in X,Y space that joins together all points that have the same quantity value.

In Figure 5.3 we have translated the highlighted information contained in the production function table into isoquants; one isoquant represents 230 units of output while the other represents 528 units of output. There are **three main points** you should note about the isoquant map we have drawn:

(i) It tells us that there are three combinations of the inputs X and Y that will enable the firm to produce 230 units of output. It tells us the same thing for 528 units of output. These alternative combinations of the inputs can be described as **alternative techniques of production**. If we let input X denote quantity of labour and input Y quantity of capital then point *a* is an example of a **capital intensive** production technique.

A capital intensive production technique is one where the quantity of capital used is relatively high per unit of labour employed.

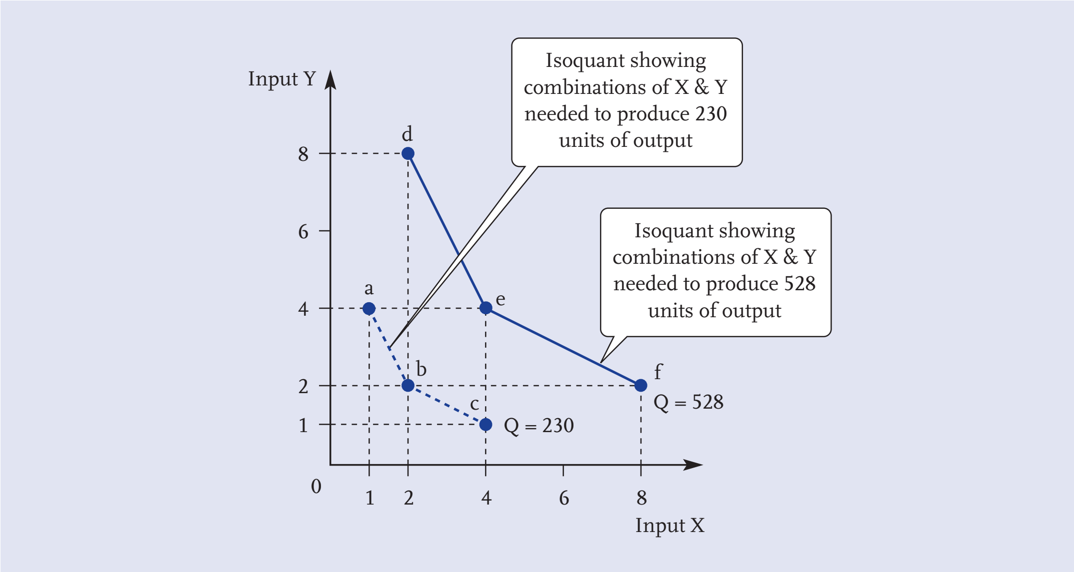


Figure 5.3 An isoquant map showing increasing returns to scale

In contrast, point *c* is an example of a **labour intensive** production technique.

A labour intensive production technique is one where the quantity of capital used is relatively low per unit of labour employed.

(ii) It is showing the presence of **increasing returns to scale**. This is easy to see because at a point such as *a* the firm can produce 230 units of output if it uses 1 unit of input X and 4 units of input Y. If the firm **doubles** the quantity of both inputs (so it uses 2 units of X and 8 units of Y) it moves to point *d* where output is 528 units, which is **more than double** that at point *a*. A similar story holds true for the pairs of points (*b*,*e*) and (*c*,*f*), respectively.

(iii) There is a **trade-off** between the inputs, which simply means that the firm can **substitute** X for Y, and vice versa. However, the **non-linear** shape of the isoquants indicates that the rate at which the firm can substitute one input for the other will vary along their length. In other words, the two inputs are **imperfect substitutes**. The rate of substitution of one input for the other is found by calculating the **slope** of the isoquant in the intervals between the points that denote the different techniques of production. We will use the Q = 230 isoquant to demonstrate.

If we move along isoquant Q = 230 from point *a* to point *b* then in order to **maintain output** at 230 units the reduction in input Y of 2 units has to be compensated for by an increase in the use of input X by 1 unit. However, if we move along isoquant Q = 230 from *b* to *c* in order to maintain output at 230 units the reduction in input Y of 1 unit has to be compensated for by an increase in the use of input X of 2 units. The ratio of the change in Y (denoted in shorthand as ΔY) to the change in X (ΔX) is called the **marginal rate of technical substitution of X for Y ($MRTS_{XY}$)**. Moving from *a* to *b* the $MRTS_{XY} = \Delta Y/\Delta X = -2/1 = 2$ (ignore the minus sign), which tells us that in the interval *a* to *b* every 2 units of Y that we use less of can be compensated for by increasing use of X by 1 unit. In the interval *b* to *c* the $MRTS_{XY} = 1/2 = \frac{1}{2}$, so we note that as we move down the isoquant from *a* through *b* to *c* the **MRTS is diminishing**. Note that if we are originally producing at *b* the fact that $MRTS_{XY} = \frac{1}{2}$ should not be taken too literally; it does not mean that we can reduce input Y from 2 units to $1\frac{1}{2}$ units while increasing X from 2 to 3 units to maintain output at 230 units because that would imply that a fourth technique of production existed and we have already stated that there are only three.

The $MRTS_{XY}$ reflects the **marginal products** of X and Y, which can be illustrated by inspection of Table 5.3 and Figure 5.3. Consider the move from *a* to *b* on isoquant 230. We can break this move down into two steps: **step 1** is the reduction in input Y while X is left unchanged; and **step 2** is the increase in input X required to compensate for the reduction in the use of input Y.

Step 1: the move from *a* to *b* consists of a reduction in input Y of 2 units. If we do not increase our use of X we will now produce 152 units of output (i.e. cell 1X, 2Y in Table 5.3). Recall that marginal product records the change in output caused by a change in the variable input when other inputs remain fixed in quantity, so if we treat X as a fixed input, then $MP_Y = (152 - 230)/-2 = -78/-2 = 39$. **Step 2**: now if we increase input X from 1 to 2 units while holding Y constant at its reduced level of 2 units we discover that $MP_X = (230 - 152)/1 = 78/1 = 78$.

You will recall from our earlier calculation that between *a* and *b* $MRTS_{XY} = \Delta Y/\Delta X = 2$, and you can now see that $MRTS_{XY}$ is also equal to the ratio $MP_X/MP_Y = 78/39 = 2$. So, more generally we can state that:

$$MRTS_{XY} = \Delta Y/\Delta X = MP_X/MP_Y$$

Now that we have identified an isoquant map and looked at some of its properties we can move on to the next stage of long-run analysis. Imagine you are an entrepreneur and you are confident that demand for your product will be 230 units per week. The long-run production function tells you that there are three techniques of production that will enable you to produce 230 units so **your problem is to choose one of the three techniques**. Note, if input Y is capital as in the Tables 4U example above, then the choice of technique is effectively a choice about the size of your production facility (often called ‘plant size’). Clearly you will want to choose the technique that allows you to produce 230 units at the **lowest possible cost**. The total costs you will incur depend on the relative input prices you will have to pay. To illustrate we will assume that the price of one unit of input X, P_X = £20 and the price of one unit of input Y, P_Y = £10. Given these input prices we can work out total production costs for the three available techniques:

- Technique *a*: total production cost = 1 × £20 + 4 × £10 = £60
- Technique *b*: total production cost = 2 × £20 + 2 × £10 = £60
- Technique *c*: total production cost = 4 × £20 + 1 × £10 = £90

Given these figures, it is rational to choose either technique *a* or technique *b*. However, **if input prices change**, these calculations will have to be repeated to find out if alternative techniques are preferable. For example, if P_X = £10 and P_Y = £20 the total costs for the three alternative techniques will now be:

- Technique *a*: total production cost = 1 × £10 + 4 × £20 = £90
- Technique *b*: total production cost = 2 × £10 + 2 × £20 = £60
- Technique *c*: total production cost = 4 × £10 + 1 × £20 = £60

Given the new relative prices it is now rational to choose either technique *b* or *c*.

We can represent the relative prices of the two inputs by drawing an isocost line in X,Y space.

An isocost line plots the combinations of the inputs that can be purchased at current relative prices for a given level of expenditure.

The **slope** of an isocost line is given by the ratio $\mathbf{-P_X/P_Y}$. So for our first example above (P_X = £20 and P_Y = £10) the slope of the isocost line will be –2. In Figure 5.4 we have plotted two isocost lines with this slope. The line closest to the origin of the graph represents a total expenditure of £60 and the line furthest from the origin a total expenditure of £90.

We can superimpose the isocost map of Figure 5.4 onto the isoquant map illustrated in Figure 5.3 to give us Figure 5.5, which shows that the rational entrepreneur will choose either technique *a* or *b* but not *c*. If *c* were chosen it would lie on a higher isocost line than *a* or *b* and as a result the entrepreneur would unnecessarily incur higher costs. If you inspect Figure 5.5 you will notice that between *a* and *b* the slope of the isoquant is identical to the slope of the isocost line. This illustrates a more general rule which states that the **optimal combination of factor inputs** to produce a given level of output is found when $MP_X/MP_Y = P_X/P_Y$, which we can rearrange as:

$$MP_X/P_X = MP_Y/P_Y$$

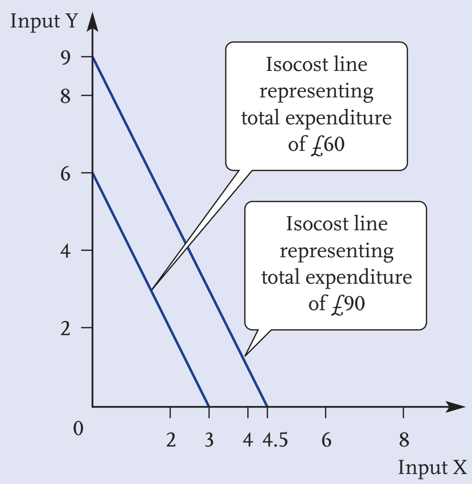


Figure 5.4 An isocost map for P_x = £20, P_Y = £10

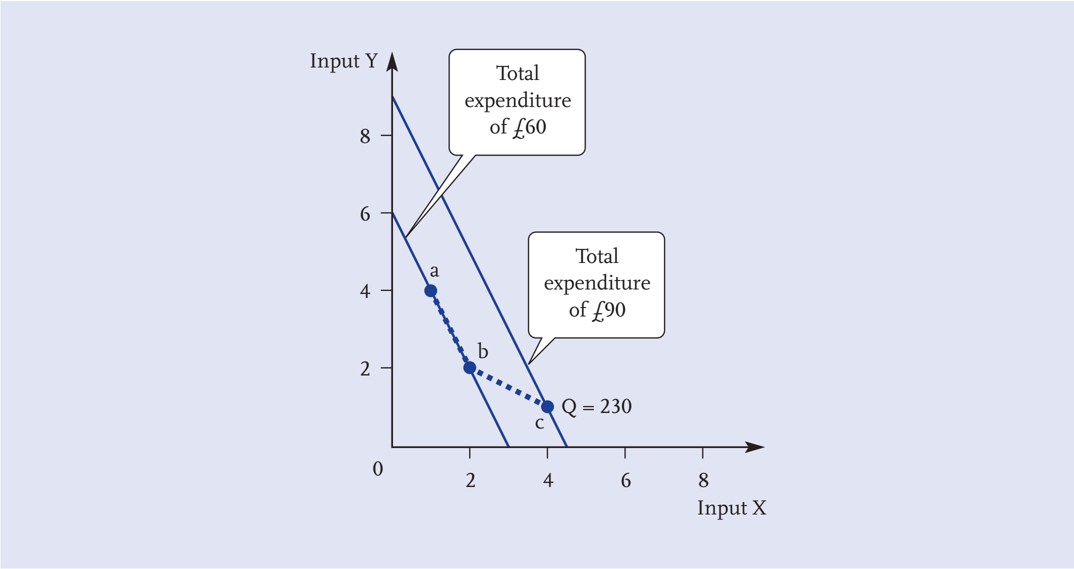


Figure 5.5 The isocost–isoquant map showing least cost techniques of production to produce 230 units of output when P_X = £20, P_Y = £10

In plain English this equation states that two inputs are optimally combined when the marginal product of the one in relation to its price is equal to the marginal product of the other in relation to its price. At point *c* this rule is violated when P_X = £20 and P_Y = £10. However, if the input prices change to P_X = £10 and P_Y = £20 the rule is satisfied at point *c* (and *b*) and diagrammatically we will observe the isocost–isoquant map shown in Figure 5.6.

Isoquants can take on a variety of shapes. We have been using discrete isoquants up to this point. This means they consist of a series of linear segments, which in turn implies that a finite number of production techniques are available. When economists build conceptual models, however, it is more common for them to **assume that isoquants are continuous**. This makes very little difference to the outcome of the analysis but it will enable a single best technique of production to be identified (as opposed to the two we have arrived at in our examples above) because an isocost line will only touch the relevant isoquant at one point. In addition, the assumption of continuous isoquants has the virtue, from the perspective of the economic model builder, of rendering the production problem amenable to the mathematical technique of differential calculus (we will not use calculus here).

We have shown an example of a production function with continuous isoquants in Figure 5.7. In our example, as we move away from the origin the distance between successive isoquants is not constant; initially the **isoquants move**

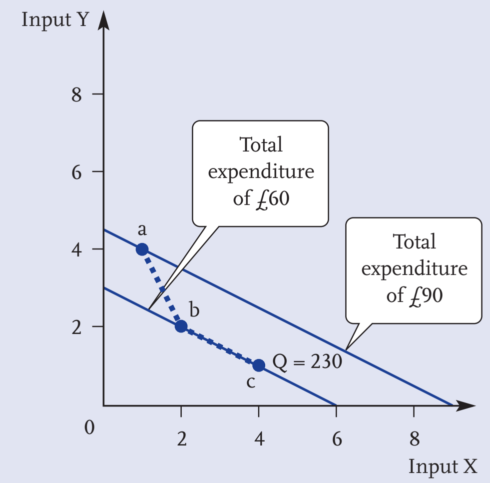


Figure 5.6 The isocost–isoquant map showing least cost techniques of production to produce 230 units of output when P_X = £10, P_Y = £20

closer together, indicating the presence of **increasing returns to scale** in the range 0 to 300 units of output.

> **Increasing returns to scale occur when the rate at which the quantity of output increases is *greater than* the rate at which the quantities of factor inputs increase.**

In the range 300 to 500 units of output the successive **isoquants are equal distances** apart, indicating the presence of **constant returns to scale.**

> **Constant returns to scale occur when the rate at which the quantity of output increases is *equal to* the rate at which the quantities of factor inputs increase.**

Finally, beyond 500 units of output the **isoquants spread out**, indicating **decreasing returns to scale.**

> **Decreasing returns to scale occur when the rate at which the quantity of output increases is *less than* the rate at which the quantities of factor inputs increase.**

With given input prices represented by the isocost lines shown in Figure 5.7 we can identify the least cost technique of production for each level of output. These are represented by the points labelled *a* through g, which trace the efficient firm's **expansion path**.

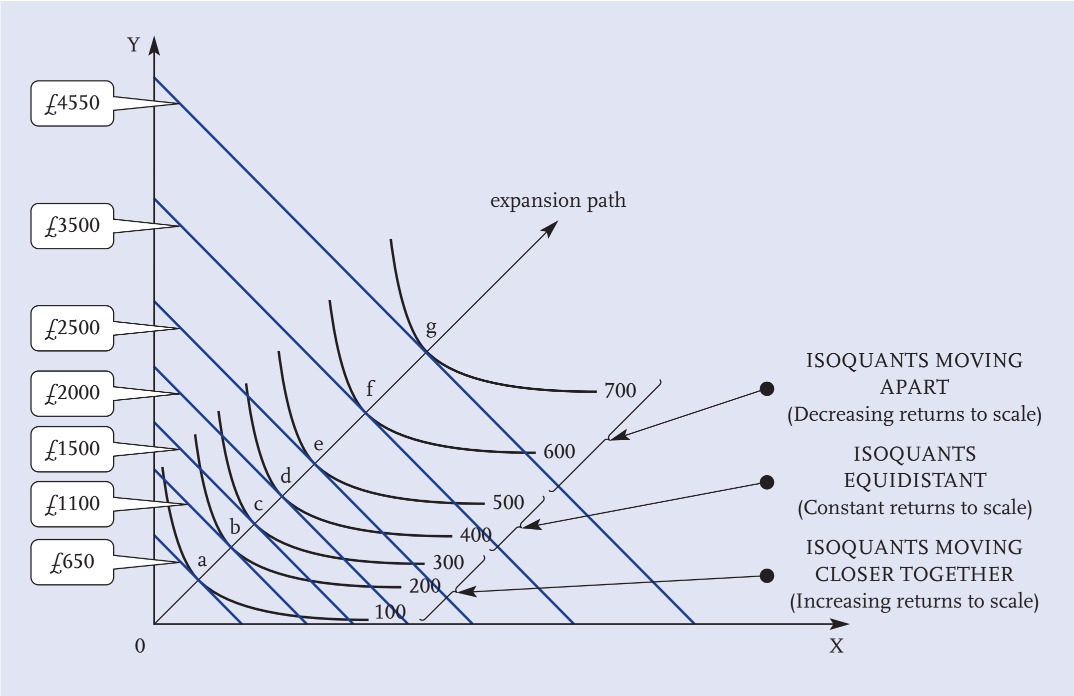


Figure 5.7 Isoquant–isocost map showing expansion path of the efficient firm

An entrepreneur faced with expanding demand for his or her product would be wise to identify the point at which decreasing returns to scale begin to take effect. If demand grows such that successively larger production facilities have to be acquired (as in the Dyson story) there may come a point where **increasing returns to scale** become **exhausted**. If this occurs then the best strategy for the entrepreneur is to build **more than one production facility** so that each facility can enjoy increasing returns to scale.

It should not have escaped your attention that we have introduced the concept of returns to scale without providing an explanation of how increasing, constant and decreasing returns occur in practical terms. Unfortunately, the mainstream approach does not deal with this question in much detail, in fact it tends to conflate it with the concept of **economies of scale** which you will meet in the next chapter. With this in mind we shall explore the theory of production further under the title of the next section in which we will get to grips with the sources of returns to scale and explore the heterodox contributions to the analysis of production.

5.4 The heterodox approach to production

5.4.1 Overview

We can view the heterodox approach to the problem of production that faces the entrepreneur as putting the flesh of detailed explanation on to the bare bones of the mainstream theory outlined above. In particular it asks questions about **how** an entrepreneur actually obtains the knowledge necessary to begin producing a good or service. This question is not asked in the mainstream theory. It is not asked because it is assumed to have been answered prior to mainstream theory taking up the story. This is equivalent to assuming that all entrepreneurs who want to enter a particular business have free access to the same detailed production function. It is a bit like putting the cart before the horse and has serious implications for our understanding of why some firms are more successful than others. There is little doubt that the production function is a useful conceptual tool, but by identifying it as the starting point for its analysis of firms the mainstream approach assumes away most of the difficult questions about production that a boundedly rational entrepreneur will have to face.

In essence, the heterodox approach can be thought of as starting with a **long-run production function** in which the cells are **blank** and asking how the entrepreneur sets about filling in the cells in a world where knowledge is less than perfect, where uncertainty is common and where learning takes place.

5.4.2 Knowledge

Having knowledge of a production function is just like having knowledge of a list of **ingredients** required to bake, say, meringues. If this is the only information you have at your disposal you will find the prospect of producing decent (edible) meringues very challenging indeed. In other words, on its own, this knowledge is of **limited value**. To make use of it you will need **complementary knowledge**. In other words, you will need a **suitable recipe**. A recipe is a set of instructions that tells you the **sequence** in which ingredients from your list are to be mixed, how long the ingredients must be cooked for and what to expect if you follow the prescribed steps. Even if you have a suitable recipe you are still likely to face a series of practical challenges, such as how to obtain usable egg whites, when you attempt to make your meringues. It is also likely that your first attempt will not be as good as your second attempt, in other words the more practice you get at making meringues the more **skilful** you will become. The point we are making here is that possession of a list of ingredients in isolation from other knowledge (i.e. a recipe, skill in applying the recipe, etc.) is not sufficient to enable you to produce successful meringues. More generally, we can say the same thing about any production function – **a production function does not encapsulate all of the knowledge that is necessary to successfully produce a good or service**.

The implication of these observations is that **different types of knowledge** exist in the world. Consequently it is extremely useful to be able to identify and classify the different types, because this will help us to understand the production problem facing an entrepreneur in greater detail. Fortunately heterodox economists have devoted a considerable amount of time to the issue of knowledge and have identified **three broad categories**. We will use the meringue example to introduce each category.

The list of ingredients for meringues is very short:

- two large egg whites
- 120 grams of caster sugar.

These ingredients are enough to make ten meringues of 6 cm diameter. This is the cook's equivalent of the knowledge imparted by a production function. We have obtained this knowledge from a cookery book. Heterodox economists call this category of knowledge **'know-that'** because it provides a list of **objective data**.

Know-that is information that can be turned into 'bits' and transmitted from one party to another. It can be further sub-divided into '*know-what*' (knowledge about facts such as the ingredients of a recipe, the population of a country, etc.) and '*know-why*' (knowledge about scientific principles such as how an internal combustion engine functions).

If know-that is the only kind of knowledge we have access to we are unlikely to bake very satisfactory meringues. We need to complement our 'know-that' with a recipe. A recipe is more generally called **'know-how'** by heterodox economists.

Know-how is the capability to perform a series of actions in order to achieve a desired result.

A recipe provides a set of instructions written by a practised cook which is designed to impart know-how to an aspiring cook.

As you may realize from your own cookery book assisted experiences in the kitchen, it is one thing to follow a recipe and quite another to produce results that look anything like those promised in the book! This is because, unlike know-that, know-how is not always easily transferred from one party to another. In our meringue example this might be because the writer of the cookery book faces difficulties in trying to convey to the reader some of the more subtle elements of their own hard won know-how. For example, meringues require the cook to obtain egg whites. This means that the yolk of the eggs and the white of the eggs have to be separated – this is not a task easily accomplished by the novice cook. However, through a more eloquent command of language some cookery book writers may be able to convey something of the subtlety of their own **experience-based skill (capability)** better than other writers are able to. Compare these two passages:

> *Tap the egg against the rim of a bowl to crack it around the middle. Holding the egg over the bowl, carefully open the shell with your thumbs, holding the two halves together to let some of the white run out. Gently tip the yolk from one half of the shell to the other, letting the white run into the bowl and taking care not to break the yolk.*
>
> (*Cooking Basics,* Hamlyn Publishers, 1999: 22)

> *Everyone always tells you that the best way to do this [separate egg whites from yolks] is by cracking open the egg and, using the broken half shells to cup the yolk, passing it from one to the other and back again ...I don't think so. All you need is for a little sharp bit of the cracked-open shell to pierce the yolk and the deal's off. It's easier and less fiddly altogether just to crack the egg over a bowl and slip the insides from their shell into the palm of your hand near the bottoms of your fingers. Then splay your fingers a fraction. The egg white will run out and drip through the cracks between your fingers into the bowl...*
>
> (*How to Eat,* Nigella Lawson, 1998: 18)

The second passage describes a different method to the generally received wisdom outlined in the first passage. Nigella Lawson has developed her method after obtaining bad results with the first method and by articulating this clearly (she certainly has an above-average dexterity with the written word) has enabled the novice cook to benefit from her own costly experiences. Nonetheless, despite the clarity of Nigella Lawson's exposition it might take a few attempts before aspiring cooks can **develop this capability** themselves, and even then the other elements of the meringue recipe will have to be tackled satisfactorily too if the end product is to be close to that desired. It is no coincidence that trainee-chefs do not learn their craft solely from books but through a series of experience-based apprenticeships under the knowing eye of a seasoned professional chef.

When knowledge is difficult to convey to third parties, either because it is impossible to articulate clearly the steps that give rise to the necessary capabilities or because the owner of the capabilities is simply unaware of the more subtle skills they have developed, it is called **tacit knowledge**.

Knowledge is tacit when it is difficult or impossible for the person who has it to articulate it clearly to a third party.

Tacit knowledge is acquired and developed entirely from experience or **learning-by-doing**. This does not necessarily mean that it is impossible to gain access to tacit knowledge, or knowledge of any other kind, that you do not possess yourself. If you possess the third category of knowledge called '**know-who**', you may be able to identify and obtain the services of a person who does possess the knowledge that you would like to use.

Know-who is the possession of information about other people or groups of people (organizations) who have knowledge that you do not possess yourself.

For convenience we have shown the categories of knowledge discussed here in Figure 5.8.

If you take another look at the Dyson story you should be able to recognize the different categories of knowledge at various points. When Dyson first recognized the original filtering problem at the Ballbarrow factory he used his own stock of *know-who* to identify the filter supplier as a possible source of *know-that* (more particularly their *know-what*) with respect to finding a solution. Once they explained to him that sawmills used a cyclonic filter he set about exploiting his own design *know-how* in order to produce a working prototype. Having done this for the Ballbarrow factory, and made the connection between this large-scale version of the technology and the small-scale technology of the vacuum cleaner filter problem, he improved his own *know-that*, in particular his knowledge of the scientific principles behind cyclonic filtration (*know-why*), and through the production of hundreds of model cyclones his *know-how* reached a high level of sophistication.

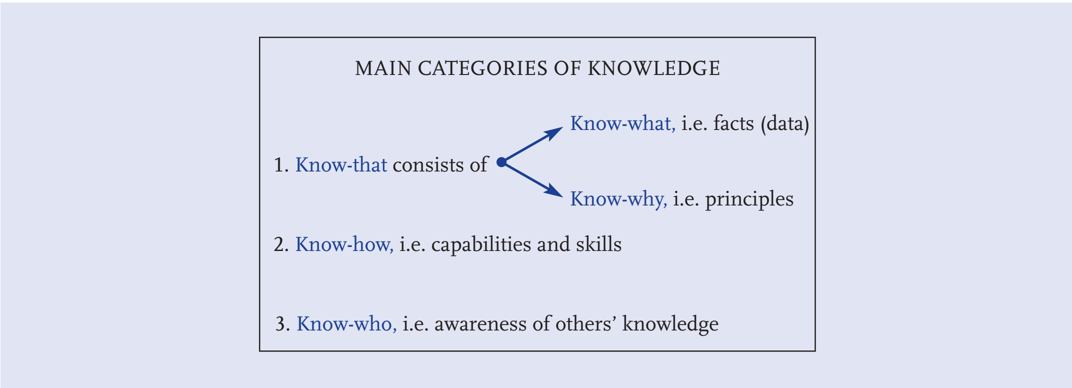


Figure 5.8 Categories of knowledge identified by heterodox economists

When it came to producing commercial products from his original designs (Cyclon and G-Force) Dyson exploited the manufacturing, distribution and marketing *know-how* of other companies through the use of licensing agreements. He learned much from this experience (one could suggest that this was his apprenticeship) and in turn developed his own manufacturing, distribution and marketing *know-how* and *know-who*. He put this to good use when he and his team (who brought in extra design and engineering *know-how*) developed the DC01 so as to minimize the number of components needed to construct it (and therefore the number of stages in the assembly process).

Dyson made use of his manufacturing and distribution *know-how* and *know-who* when he opened his own vacuum cleaner assembly facility (which required him to find reliable suppliers of inputs, to find willing retail stockists and to organize the assembly line). In the original Dyson factory at Chippenham it took a while for the newly created workforce to acquaint themselves with the production facilities but within two weeks they had developed their assembly *know-how* sufficiently to be able to produce 100 vacuum cleaners a day. Before too long Dyson was exploiting his hard won know-how even further when he introduced the DC02 cylinder-style vacuum cleaner.

It should be clear from our discussion here that knowledge comes in various guises and from a variety of sources and that the capability to produce something is not simply a matter of access to *know-that*. It should also be clear that entrepreneurs need to recognize this and **actively develop** their **knowledge** in such a way that they can turn business ideas into practical realities. To be successful this process will require the exploitation and development of all of the categories of knowledge we have identified and it is likely to take a significant amount of **time** during which a lot of **learning** will take place. Bearing these thoughts in mind we will turn our focus of attention to the concept of returns to scale.

5.4.3 Returns to scale

In the production function illustrated in Figure 5.7 we observe increasing returns to scale followed by constant returns to scale and eventually decreasing returns to scale as output quantity is increased. How do economists explain these phenomena? We will begin our analysis with a visit to Samantha Pinewood's evolving business.

It is now several months since Tables 4U began production of its flat-packs for sale to Pinewørld. The flat-packs are proving to be very popular with Pinewørld's customers so they decide to increase their weekly order from 450 units to 900 units. If Samantha is going to be able to meet this order she will have to expand her production facilities because, as we have seen, with just three units of capital, diminishing returns set in at relatively low levels of output and the capacity of the workshop peaks at 600 flat-packs. The problem Samantha faces is how best to produce 900 flat-packs. In order to get to grips with this long-run problem Samantha draws upon the specialized knowledge of an engineer whom she asks to provide an estimate of how many units of capital she should invest in in order to produce 900 flat-packs. The first thing the engineer does is take a look at the **rate of use** of Tables 4U's current capital equipment. This reveals something interesting; the lathe and the bandsaw have both been used 100 per cent of the time during the last few months of production but the flat-packing machine has been lying idle for 50 per cent of the time. This information indicates that the flat-pack machine is underutilized at the present rate of output; that is, it is capable of easily handling the volume of output provided by the solitary lathe and the solitary bandsaw. Consequently the engineer estimates that Samantha needs to invest in another lathe and another bandsaw but not another packing machine. To accommodate the two extra units of capital will mean extending the workshop to 500 square feet. He provides an estimate of the short-run production function for a 500 square-foot workshop which will apply *if Samantha organizes her capital and labour 'appropriately'*. This is shown in Table 5.4.

Samantha inspects the data gathered from her original workshop (Table 5.1) and compares them with the data provided by the engineer for the larger workshop (Table 5.4). She notes from Table 5.1 that with her current workshop she is able to produce 450 flat-packs by employing three units of capital and three units of labour, while Table 5.4 tells her that she will be able to double this level of output to 900 flat-packs if she increases her factor inputs to five units of capital and five units of labour. In other words, expansion of the production facilities will enable Tables 4U to enjoy **increasing returns to scale** (that is, it can double its output without the need to double its factor inputs). How is this possible? The answer is that at a higher level of output Tables 4U can make much better use of the packing machine which was underutilized in the smaller workshop. Even though the packing machine was underutilized in the smaller workshop, Tables 4U could not dispense with its services altogether. In such a situation capital is said to be **indivisible**.

Capital is indivisible when it cannot be obtained in quantities that provide a rate of service that exactly matches the firm's requirements. This simply means that the firm cannot obtain a fraction, say half, of a packing machine. Instead a whole packing machine has to be acquired and used only half as often as it is capable of being used.

Table 5.4 Weekly production of flat-packs at Tables 4U with varying levels of labour usage (500 sq. ft. workshop housing five units of capital)

Quantity of capital (K)	No. of workers employed (L)	Number of flat-packs per week (TP)	Marginal product of labour (MP_L)	Average product of labour (AP_L)
5	1	100		100
			150	
5	2	250		125
			200	
5	3	450		150
			200	
5	4	650		163
			250	
5	5	900		180
			250	
5	6	1150		192
			200	
5	7	1350		193
			100	
5	8	1450		181
			−150	
5	9	1300		144

Increasing returns to scale in our example arise because at larger outputs the firm is able to overcome the indivisibility problem. In other words, even though another lathe and another saw have to be obtained to enable Tables 4U to meet the new quantity ordered by Pinewørld the firm does not have to obtain another packing machine.

You should note that our discussion about the sources of increasing returns to scale has assumed that Samantha is capable of organizing the capital and labour at her disposal 'appropriately'. We made the same assumption when we introduced the short-run production function for three units of capital shown in Table 5.1 where we also briefly introduced the related concept of the **division of labour**. We will now examine this concept in greater detail.

The division of labour is the practice of dividing the production process into its component tasks and allocating each one of these tasks to a particular member of the labour force. Such a unit of labour is said to have *specialized* in that specific task.

The division of labour was first introduced to economics by Adam Smith over 200 years ago in his famous book *An Inquiry into the Nature and Causes of the Wealth of Nations*. Mainstream economics has taken Smith's message on board but it does not provide any detail about how the division of labour should be achieved to best exploit its benefits. Instead, as we have seen, a mainstream production function simply relates a quantity of inputs to a quantity of output without saying how the tasks involved should be organized and co-ordinated. The significance of this point is that the successful division of labour requires the entrepreneur to possess some **organizational know-how**. This is for the very simple reason that the division of labour is an **organizational concept** that relies upon the entrepreneur being able to organize the **flow of work** between the different stages of the production process adequately. Furthermore, it provides us with another very important reason for increasing returns to scale.

Adam Smith's original analysis of the division of labour used the example of pin production. He pointed out that pins, or anything else for that matter, can be produced either by a single multi-skilled person (an artisan) who carries out all of the stages of production or, alternatively, each stage of production can be allocated to different members of the labour force. Here's what he said:

> *...a workman not educated to this business [pin manufacture]...nor acquainted with the use of machinery employed in it...could scarce, perhaps, with his utmost industry make one pin a day and certainly could not make twenty. But in the way in which the business is now carried on...it is divided into a number of branches...One man draws out the wire, another straights it, a third cuts it, a fourth points it, a fifth grinds it at the top for receiving the head; to make the head requires two or three distinct operations; to put it on, is a peculiar business, to whiten the pins is another; it is even a trade by itself to put them into the paper; and the important business of making a pin is, in this manner, divided into about eighteen distinct operations, which, in some manufactories, are all performed by distinct hands, though in others the same man will sometimes perform two or three of them. I have seen a small manufactory of this kind where ten men only were employed...they could, when they exerted themselves, make among them...forty-eight thousand pins in a day...if they had all wrought separately and independently...they certainly could not each of them have made twenty...*
>
> Adam Smith (1776/1986: 109–10)

The implications of Smith's analysis are far reaching. In simple terms, one major **implication** is that the greater the number of stages a production process contains then the greater the scope for dividing labour becomes. However, it is pointless employing greater quantities of labour and allocating each person to a specific task if there is insufficient **demand for the product** to support larger scale production facilities. So if the size, or **extent**, of the market for the product is **limited** then so is the scope for further division of labour. As Smith put it:

> *As it is the power of exchanging that gives occasion to the division of labour, so the extent of this division must always be limited by the extent of that power, or in other words, by the extent of the market.*
>
> ibid, 121

But why should dividing labour and allocating each unit to a specialized task lead to increasing returns to scale? Smith's answer rests on **three propositions**:

(i) Workers who specialize in performing only one task on a production line will experience an improvement in their capability to perform that particular job that is greater than if they are expected to perform a multiple set of tasks. Clearly a learning process is implied here:

> *The different operations into which the making of a pin...is subdivided, are all of them much more simple, and the dexterity of the person, of whose life it has been the sole business to perform them, is usually much greater. The rapidity with which some of the operations of those manufacturers are performed, exceeds what the human hand could, by those who had never seen them, be supposed capable of acquiring.*
>
> ibid, 113

(ii) If workers do not have to move from one task to another (e.g. from one workbench to another in the Tables 4U example) then a considerable amount of time will be saved. Smith points out that time is also saved by the fact that workers who do not have to switch between different tasks, each of which requires different skills, will not have to use their brains in order to think about the new task at hand:

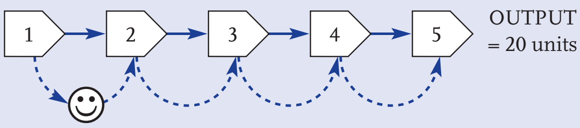
> *It is impossible to pass very quickly from one kind of work to another that is carried on in a different place and with quite different tools...*
>
> ibid, 113

(iii) As workers become more practised at specialized tasks they are likely to adapt their tools so they help them to become even more efficient at the task:

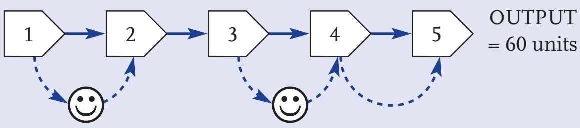
> *Men are much more likely to discover easier and readier methods of attaining any object, when the whole attention of their minds is directed towards that single object, than when it is dissipated among a great variety of things.*
>
> ibid, 114

Assumptions:

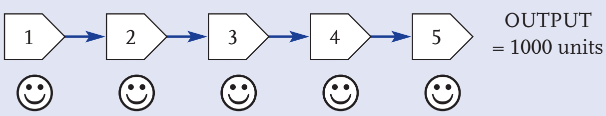
(1) Production takes place over 5 successive stages.
(2) Each stage uses a specialized piece of capital.
(3) Each piece of capital requires a human operator.
(4) The capital at stages 3 and 4 is capable of operating at a rate that is at least twice that of the capital at the other stages.



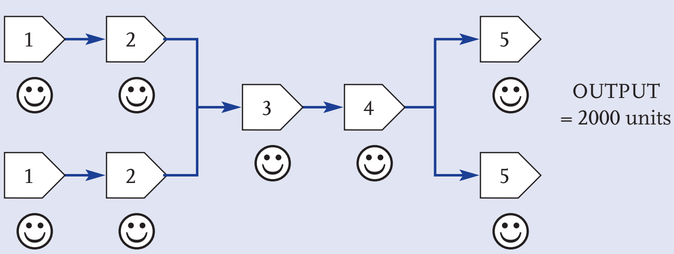
Case 1: when size of market is 'small' a single unit of labour, who works at his/her own pace, operates capital at all successive stages of production. Labour productivity is very low.



Case 2: when size of market is slightly bigger than case 1, two units of labour, who adjust their own work rates in order to synchronize the successive stages of production to avoid stockpiling and lags, operate a subset of capital each. This is PARTIAL DIVISION OF LABOUR. Labour productivity is enhanced modestly.



Case 3: size of market has grown sufficiently to allow one unit of labour per unit of capital. This is FULL DIVISION OF LABOUR but the labour at stages 3 and 4 are not working at the same rate as the labour at stages 1, 2 and 5 because of the rate at which 'their' respective units of capital are capable of working. Labour productivity is significantly increased.



Case 4: when size of market is 'large' capital can be duplicated at stages 1, 2 and 5 and at stages 3 and 4. INDIVISIBILITY OF CAPITAL IS OVERCOME and accompanied by FULL DIVISION OF LABOUR. Labour productivity is enhanced once again.

Key:

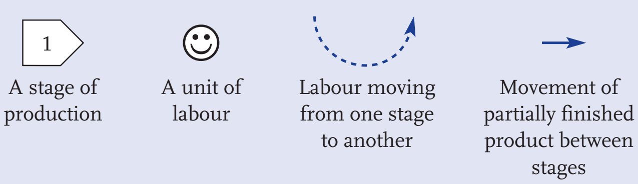


Figure 5.9 An example of increasing returns to scale due to exploitation of the division of labour and overcoming capital indivisibility as the extent of the market grows

To help us clarify the reasons for increasing returns to scale Figure 5.9 illustrates a simple example which shows how an entrepreneur can organize the firm's production process in **four different configurations** as the **size of the market** for its product **expands**.

As we have seen, Adam Smith's own estimates of the benefits from the division of labour in the pin factory example provide some quite startling increases in productivity and output (one man can produce 20 pins, while ten men can produce 48 000 pins in a day). Other writers have pointed to similarly startling productivity improvements in studies of their own. A particularly famous example is the evolution of the production line for the Ford Model T as described by Alfred Chandler in his book, *The Visible Hand*:

> *Ford and his colleagues adopted the most advanced machinery,... and followed the 'line production system' of placing machines and their operators in a carefully planned sequence of operations. Ford's factory engineers designed improved conveyors, rollways, and gravity slides to assure a continuing regular flow of materials into the plant. These engineers also began to experiment with the use of conveyor belts to move parts past the worker doing the assembly, with each man assigned a single highly specialized task. The moving line was first tried in assembling the flywheel magneto, then other parts of the engine, next the engine itself, and finally, in October 1913, in assembling the chassis and the completed car. The innovation – the moving assembly line – was an immediate success. The speed of throughput soared. Labor [sic] time expended in making a model T dropped from 12 hours and 8 minutes to 2 hours and 35 minutes per car. By the spring of 1914...the average labor time per car dropped to 1 hour and 33 minutes.*
>
> Alfred Chandler (1977, p. 280)

So, increasing returns to scale can be enjoyed by increasing the division of labour and also overcoming indivisibilities in capital equipment as the extent of the market grows, but there is a third reason too. If you look carefully at Figure 5.9 you will note that we have made the explicit assumption that capital equipment is specialized to the particular task it is assigned in the production sequence. The division of labour therefore involves each unit of labour becoming an expert, that is a specialist, in the use of a particular piece of specialized capital equipment. There is no reason, however, that further increasing returns to scale cannot be enjoyed by the firm if the entrepreneur is able to **automate several specialized tasks** so that they are performed not by a series of separate pieces of capital but instead by one **integrated piece of capital**. If this is possible, the firm can reduce its quantity of capital and, if the integrated piece of capital can be attended by a single human operator, its quantity of labour can be reduced also. Some writers have suggested that this is exactly the process that has occurred in the late twentieth and early twenty-first centuries in the manufacturing sectors of the developed economies of the world, and they have used this argument to explain why a significant proportion of the labour force in these economies is increasingly employed in service industries.

We will finish this section by explaining the reasons we might observe constant returns to scale and decreasing returns to scale. Firstly, **constant returns to scale** will occur when all of the opportunities for further division of labour have been exhausted and when all indivisibilities have been overcome; however, they will only persist as long as the reasons for decreasing returns to scale have not taken hold.

To understand why **decreasing returns to scale** arise we need to raise three points. The first point is that an entrepreneur has to make a conscious effort to organize the factors of production. This is not a trivial task. The second point is that the entrepreneur will have to oversee the production process and manage the people involved (an example of a management task is making sure that all workers are using their best efforts rather than shirking on the job). The third point is that the entrepreneur is, like all human beings, boundedly rational with all that this implies with respect to his or her ability to be able to cope with an increasing workload. The first two points in combination with the third imply that beyond a certain size of organization the entrepreneur will be unable to give the firm the degree of effective management it requires. So, in rather blunt terms, decreasing returns to scale may arise because the entrepreneur **loses effective control** of the firm due to his or her **cognitive limitations** coupled with the increased size and **complexity** of the production process. Given that different entrepreneurs are endowed with different degrees of cognitive ability one may be able to manage effectively a given size of firm where another may experience a degree of control loss. As you will discover later in the book, many firms have discovered ways of overcoming this problem to a greater or lesser extent, but these 'solutions' have brought their own problems.

5.5 Summary

We have covered a lot of ground in this chapter and along the way we have introduced you to a number of fundamentally important concepts. At first reading you may find it difficult to make sense of all of the arguments we have outlined, especially as some of them seem to conflict with each other. Do not worry about this – it is perfectly normal; even professional economists have trouble keeping tabs on all of the strands!

In order to help you make better sense of everything we have discussed you might find it beneficial to re-read the details of the Dyson story in the light of the theoretical discussions that have taken place in the rest of the chapter. Given that the heterodox approach takes as its point of departure a critical look at the mainstream production function approach, in particular the assumptions it makes about knowledge and organization, you may be tempted to think that the mainstream approach is of little practical use to you. This would, however, be a hasty conclusion to jump to. It is very rarely a good idea to ignore the lessons provided by 100 years of theoretical development! Yes, the production function approach does make some rather strong assumptions with respect to the nature and availability of knowledge, but this is simply because the theories of mainstream economics are built upon the assumption that a human decision maker, which in this chapter is the entrepreneur, has full information about all relevant facts and is perfectly capable of processing this information optimally.

While this assumption might not accord with the world we live in, and indeed the world James Dyson occupied, it is nevertheless a useful assumption to make *as a first step* to help us frame some of the issues that are relevant to the problem of production. In particular it helps us to identify easily the concepts of diminishing marginal returns, returns to scale and capital intensive and labour intensive production techniques – all of which are concepts an entrepreneur needs to be aware of. However, if we accepted the mainstream approach unquestioningly we would provide a message for entrepreneurs that runs as follows: you and all other entrepreneurs like you have access to the latest production techniques and full knowledge of them; in addition you are all as equally capable of utilizing these techniques. As a result, if any of you start a business enterprise then you can expect to be no better at carrying out the tasks associated with production in this business than any other entrepreneur – consequently, if you wish to make decent profits you will need to create some kind of barrier to entry into the industry you are a part of. We have already

5.5 **Summay** (continued)

mentioned this in Chapter 3 and we will revisit it again in greater detail in Chapter 7. For now you can note that it is a generally recurring theme in the mainstream approach.

Once we began to look at the production problem through the eyes of a heterodox economist we dispensed with the globally rational, infinitely capable model of the entrepreneur and we started to pose questions that asked how a boundedly rational entrepreneur could obtain the knowledge that is necessary to provide the data that a mainstream production function take as given. This model of the entrepreneur is more 'realistic' in the sense that it is more like the people we see around us every day. The questions we asked led us to distinguish between three broad categories of knowledge and, in turn, these revealed that we need to distinguish between the data recorded in the cells of a production function and the ability to actually turn the relationship between quantity of inputs and quantity of output shown there into a practical reality. This requires the entrepreneur to be able to organize the factors of production appropriately.

A boundedly rational entrepreneur needs to acquire and develop not only *know-that* (which is equivalent to having knowledge of the data recorded in the cells of the production function), but also *know-how* (which is often referred to as capabilities) and in all likelihood *know-who*. Learning was therefore identified as a very important part of the heterodox analysis of the production problem, which also implied that the time taken to achieve successful production is another salient issue that needs to be taken seriously by the entrepreneur.

Finally we used the broader concept of knowledge to help us understand better the concept of returns to scale, and we showed that specialization through the division of labour is absolutely key here, along with the concept of indivisibility of capital. You should note also that, because of the pervasive influence of uncertainty, when an entrepreneur makes the initial decision regarding the scale of production facilities, he or she is unlikely to know with any degree of accuracy the likely extent of the market for the firm's product. Consequently, we might suggest that not only is the division of labour limited by the extent of the market, but it is also limited by the **predictability** of the extent of the market; as we saw, James Dyson grew the scale of his production facilities *organically* as the extent of his market grew and revealed itself to him.

The heterodox message for would-be entrepreneurs is slightly different from the mainstream message; if we accept that entrepreneurs need to

5.5 **Summary** (continued)

develop particular capabilities that are peculiar to their chosen production process, and if such *know-how* has been difficult to develop (and would be equally as difficult to explain to someone else), then maybe different but competing businesses will experience different degrees of success because of their **different levels of capability**. We will return to this point in Chapter 8.

One final point: in emphasizing learning-by-doing, the heterodox approach has some interesting implications for the growing firm's expansion path. If you look at the isoquant–isocost map shown in Figure 5.7, it should be evident to you that the locus of the expansion path followed by the firm illustrated there depends upon the **relative prices** of the factor inputs X and Y, as shown by the slopes of the isocost lines. This implies that **a change in relative prices** at any time will cause the firm to **change its technique of production** (i.e. the relative mix of quantities of X and Y). This analysis does not sit easily with the learning story told by the heterodox approach.

The mainstream approach does not sit well with a learning approach because, if knowledge is not easy to obtain, then once a particular technique has been chosen further expansion and development of the firm's production capability is likely to be based upon **learning** more about **this particular technique**. In other words, while alternative production techniques might well exist (as implied by an isoquant) the fact that the entrepreneur has made an **initial choice** may well mean that he or she devotes all of his or her (scarce) attention, (scarce) energies and (scarce) time in the future to developing the firm's capability at using this particular technique to the **exclusion of all alternatives**. The implication of this is that as time passes the entrepreneur will become more and more knowledgable about the firm's chosen technique and relatively **ignorant of alternatives**. Furthermore, the firm will have developed its capability with the technique to quite a high degree of sophistication (it might even have developed integrated labour-saving capital).

In such a situation it might be very difficult for the firm immediately to change technique in response to factor price changes. This is not to say that a change in technique is entirely out of the question, but it is likely to require the firm to undergo a period of painful and costly learning about the new alternative (and 'un-learning' of the original). These costs, which are over and above the direct costs incurred as a result of hiring the factors of production, are called **switching costs**. If switching costs are considered to be *too high* the firm will carry on with its original technique. If this happens the

5.5 **Summary** (continued)

firm is said to be **locked in** to its original production technique. In other words, the choice made by the entrepreneur at the birth of the firm has determined how it will do things later on in its life.

When the number of options for later decisions about the technique of production is restricted by the decision made in the initial period, economists say that the firm's expansion path exhibits **path dependence**. Put another way, path dependence suggests that if different entrepreneurs in the same line of business make different decisions at the birth of their firms then we might see firms **evolving** along quite different trajectories from each other. This sits in stark contrast to the mainstream analysis of production where it is assumed that firms will be able to jump easily and without cost from one technique to another in response to factor price changes, and that every entrepreneur will make identical decisions in this regard.

5.6 Some questions to consider

1. Does the phrase 'a unit of capital' have a precise meaning?
2. If you were asked to define what is meant by the phrase 'technique of production' how would you start? Is the mainstream notion of technique of production as being either capital intensive or labour intensive helpful?
3. Do you think it is likely that the quantities of output recorded in the cells of a production function will stay constant over time? Give reasons for your answer.
4. Think of some of the things you have learned in your life (including skills) and attempt to allocate each to one of the categories of knowledge identified by the heterodox approach. Can you think of any categories of knowledge that are *not* captured by the heterodox classification?
5. Do you think that the process of the division of labour means that we will see firms becoming larger and larger? What role do you see for small firms in the economy?
6. Can you think of any decisions that you have made in your life that have locked you in to a particular set of options at a later date?

5.7 Recommended additional reading sources

For the story of how some of the world's major businesses have coped with the problems of production, see Alfred Chandler (1977) *The Visible Hand: the Managerial Revolution in American Business*, Harvard, Belknap Press.

The Dyson story is told in great detail by Giles Coren (2002) *James Dyson Against the Odds: An Autobiography* (New edition), London, Texere Publishing. All of the Dyson quotations used in this chapter are taken from this book.

For an interesting view of the theory of production, see Nicolai Foss (1997) 'The classical theory of production and the capabilities view of the firm', *Journal of Economic Studies*, **24**, no. 5, pp. 307–23.

A more advanced discussion of some of the topics we have covered in this chapter is provided by Richard Langlois (1999) 'Scale, Scope and the Reuse of Knowledge', in S.C. Dow and P.E. Earl (eds) *Economic Organization and Economic Knowledge: Essays in Honour of Brian J. Loasby*, **1**, Cheltenham, Edward Elgar, pp. 239–54.

Another advanced discussion of the topics we have covered in this chapter, which is complementary to Langlois' article, is given by Axel Leijonhufvud (1986) 'Capitalism and the factory system', in R.N. Langlois (ed.) *Economics as a Process: Essays in the New Institutional Economics*, New York, Cambridge University Press, pp. 203–23.

Our discussion of types of knowledge and the production function was informed by Brian Loasby (1999) *Knowledge, Institutions and Evolution in Economics*, London, Routledge (see Chapter 4, 'Capabilities', pp. 49–68).

The original account of the division of labour can be found in Adam Smith (1776/1986) *An Inquiry into the Nature and Causes of the Wealth of Nations*, London, Penguin Books (see Book 1, Chapters I–III).