III

THE NEOCLASSICAL PERIOD, 1890–1939

14

Equilibrium Analysis

14.1 INTRODUCTION

The new systems proposed by Jevons, Menger, Walras and Marshall were all systems of static equilibrium, in which prices were determined by the interaction, in competitive markets, of the maximizing behaviour of economic agents. The emphasis was thus very different from that of the classical economists for whom a static system of price determination, though present, was in the background *vis à vis* the theory of growth and capital accumulation. It is the development of this system of static equilibrium, which though propounded in the 1870s, was far from fully worked out, that forms the subject of this chapter. It can be argued that the working out of this system was the greatest achievement of the period leading up to 1914.

There were a variety of ways in which this development could proceed, for static equilibrium theory was formulated in widely differing ways, each having its adherents. There were differences of opinion with regard to utility, the nature of costs, the usefulness of mathematics and the issue of simultaneous determination versus one-way causation. In addition, there were still, in the years after 1870, many economists who retained older approaches to the theory of value. Furthermore, despite isolated attempts, the problem of income distribution had not, before the 1890s, been systematically reformulated in the same way as the theory of commodity pricing. Though their nature changed markedly, discussions of the nature of utility, the way in which economic equilibrium should be handled, and controversies over the theory of distribution continued throughout the period.

14.2 THE CONSUMER AND DEMAND

In a sense the theory of the consumer was fundamental to all the new systems, for values were derived from the values of final consumption goods, these in turn being derived from the ability of goods to satisfy wants. Although the way in which consumer theory was expounded varied from Jevons' Benthamite utilitarianism to Menger's stress on the ability of goods to satify human needs, all these writers analysed consumer behaviour in terms of utility. It was marginal utility which determined the value of a commodity. In addition, for all of them, the utility of a commodity was assumed to depend solely on consumption of that commodity.

There were several ways in which these investigations could be developed. One was to investigate the nature of utility: could it be measured, and if so how? What was its connection with hedonism? The second way was to introduce technical improvements into utility analysis, permitting a more thorough analysis of demand. Finally, there was the approach of abandoning utility analysis in favour of alternative ways of representing consumer behaviour. In the period from 1880 to 1939 all three of these approaches were pursued.

The meaning of the term utility[1]

For economists wishing to investigate further the nature of utility, one recourse was to psychology. The only major economist of the period to follow Jevons[2] in adopting this approach was Edgeworth. Edgeworth made use of the Weber–Fechner laws of sensation, dating from the 1860s, taking as his unit of measurement the "just perceivable increment", something applicable to both the intensity and the duration of pleasure. This "equation to each other of indistinguishable events or cases" he regarded as axiomatic, incapable of proof.[3]

Edgeworth was, however, atypical, the main reason for this being that amongst psychologists interest in hedonism was waning. Several writers, of whom the first was Bonar (1888),[4] warned economists of this, raising the question of whether utility analysis had to fall with hedonism. Reactions to this amongst more orthodox economists were varied. At one extreme we have Marshall. In the first edition of the *Principles* (1890) the utilitarian basis of the theory was explicit, but in later editions, though the issue was never explicitly discussed, references to "pleasure" and "pain" were replaced with more innocuous words such as "satisfaction" and "detriment". The substance of the theory was unaffected. At the other extreme we have Veblen's attacks (1898, 1899), not only on hedonism but on the notion of rational choice itself.[5]

But of greater importance than either of these extremes were the attempts of economists to provide accounts of consumer behaviour free of hedonism, the most important contributions being those of Fisher and Pareto. Fisher (1892) described the foisting of psychology on economics by Gossen, Jevons and Edgeworth as "inappropriate and vicious", arguing that it was possible to base economics on the much simpler postulate, "each individual acts as he desires".[6] Though Fisher continued to use the word "utility" it was merely as a means of describing behaviour: to say that the utility of A exceeded that of B meant no more than that the individual preferred A to B. Referring to Bentham, Fisher argued that "his *word* [utility] is the more acceptable, the less it is entangled with his *theory*".[7] The distance between Fisher's conception of utility and that of Bentham and Jevons is shown by the following paragraph.

> Thus if we seek only the causation of the *objective facts of prices and commodity distribution* four attributes of utility as a quantity are entirely unessential, (1) that one man's utility can be compared to another's, (2) that for the same individual the

marginal utilities at one consumption-combination can be compared with those at another, or at one time with another, (3) even if they could, total utility and gain might not be integrable, (4) even if they were, there would be no need of determining the constants of integration.[8]

A similar position was taken by Pareto who, apparently independently of Fisher, noticed that a utility function might not exist. He approached this problem in the following way. Slopes of indifference curves (which will equal the slopes of the appropriate price lines) can be obtained from budgetary data. To obtain utility functions from these, two further steps are necessary: (i) to integrate equations for the slopes of the indifference curves to obtain the indifference curves themselves; and (ii) to integrate these equations to obtain utility functions. Pareto observed that whilst the first step might be possible under certain conditions, there was no way of going from the indifference curves to a unique utility function. Like Fisher, Pareto was clearly taking observed behaviour as the datum, utility functions being no more than a way of representing this. It was for this reason that he advocated the term *ophélimité* for utility. Pareto was, however, slow to see the implications of these arguments. He noted the problem as early as 1892,[9] but in the *Cours* (1896) he not only used utilities, but assumed that the utilities of different individuals could be compared. Even the *Manual* is not entirely consistent in its attitude to utility measurement. Despite this, however, Pareto must, along with Fisher, be regarded as one of the main architects of modern, ordinal, utility theory.

Pareto was in a minority amongst leading economists of the late nineteenth century in rejecting the measurability of utility. Wicksell, Wicksteed and Edgeworth all advocated measurable utility, and Marshall considered utility to be measurable under certain special circumstances. In the period after 1900, however, significant contributions were made towards developing a non-utilitarian consumer theory: Johnson (1913), apparently independently of Pareto, analysed consumer choice in terms of the ratios of marginal utilities, arguing that economics did not need to know the marginal utility of any commodity[10]; and Slutsky (1915), referring to Pareto but not Johnson, attempted to provide a completely empirical concept of utility. Their contributions were, however, neglected, and it was not until the 1930s that the concept of purely ordinal utility became generally accepted.[11] Hicks and Allen were the main proponents of this approach, whilst Lange was responsible for pointing out exactly what assumptions were involved in the various types of utility. But to understand these developments we need to consider other aspects of consumer theory, to which we now turn.

Utility and demand

When shorn of its utilitarian associations the function of utility theory was to provide a theory of demand. Both Walras and Marshall used utility maximization to derive their demand curves, but they did so only for the case where the utility of each good depended on the consumption of that

good alone, i.e. for an additively separable utility function.[12] The problem with this was that, although easy to understand, it ruled out the issue of complementarity: for the essence of complementarity is that the utility of one good depends on how much of another good is being consumed. Thus an important step was the replacement of the additively separable utility function with a more general one. The first economist to do this was Edgeworth (1881). He, however, did not pursue the link between complementarity and the form of the utility function.

The first formal definition of complementarity was provided by Auspitz and Lieben (1889) who defined it in terms of the second derivative of the utility function: if an increase in consumption of one good increases the marginal utility of another, then they are complementary; if it reduces it they are competing goods. Despite their rejection of measurable utility, on which this definition depends, both Fisher and Pareto took over this definition. It was not until Johnson's article in 1913 that a definition of complementarity that did not depend on utility measurement was provided. This defined complementarity in terms of the slope of the indifference curve. The modern definition, in terms of the sign of cross substitution effects (i.e. according to whether a rise in the price of one good increases or reduces the demand for another), was provided by Hicks and Allen in 1934.

Complementarity was not, however, the only reason why the form of the utility function mattered. It was because they assumed utility functions to be additively separable that Marshall and Walras were able to deduce that demand curves sloped downwards. When Fisher and Pareto analysed this problem using more general utility functions they found that in general demand functions might slope either way. Here again it was Johnson and Slutsky who solved the problem, the most thorough treatment being that of Slutsky. Slutsky distinguished between what we now call normal and inferior goods, calling them "relatively indispensable" and "relatively dispensable". Demand curves for the former necessarily sloped downwards; those for the latter probably sloped downwards, but did not necessarily do so. Because of the neglect of Johnson's and Slutsky's work, however, Marshall's analysis of demand continued to be used, despite its limitations. It was only when Hicks and Allen derived these results, independently of Johnson and Slutsky, though they later acknowledged them fully naming the fundamental equation after Slutsky, that the generally held theory of demand changed.

Indifference curves

These developments in consumer theory required a means of analysing consumer choice without using utility, and this was found in indifference curves. However, despite the use of indifference curves by Fisher, Pareto, Johnson, Hicks and Allen as a means of avoiding reliance on utility, indifference curves were first used by a committed utilitarian – Edgeworth. Because Edgeworth used his indifference curves to analyse bargaining, they sloped upwards, relating an individual's consumption of one good to the

amount of the other good he had to give up. It was Fisher and Pareto who saw in indifference curves the possibility of developing a non-utilitarian analysis of consumer behaviour, and who used the now conventional downward-sloping indifference curves. Fisher combined these with a budget line,[13] whilst Pareto combined them with constraints which might, or might not be straight. It was Johnson who used indifference curves and budget lines to derive the now familiar income- and price-consumption curves.[14]

As with other aspects of consumer theory, Hicks and Allen re-worked much of this, taking the analysis further. One key to their treatment was the concept of the elasticity of substitution, developed by Hicks, Robinson and Lerner to analyse the production function, but used by Hicks and Allen to measure the curvature of indifference curves.[15] Their contribution, no doubt a major reason for the success of their work, was to show that indifference curves provided a way of analysing demand without making any assumptions about the effects of income on demand. One of the problems with Marshall's theory of demand was that he had had to assume that the marginal utility of money was constant. It was this assumption which had enabled Marshall to measure marginal utility in terms of money, and to infer a downward-sloping demand curve from the hypothesis of diminishing marginal utility.[16] Fisher and Pareto had shown that indifference curves could be used to analyse demand but without showing that this made it possible to dispense with the Marshallian assumption. It was Hicks and Allen who did this.

More radical approaches

The use of indifference curves and ordinal utility indices enabled economists to dispense with utilitarianism. There were, however, economists who wished to go even further than this. One such was Cassel (1899, 1918a) who advocated a return to Cournot's approach of starting with demand functions. He argued that it was sufficient to assume "that the demand for each of the articles in question is settled as soon as the prices of these articles are fixed."[17] To analyse economic problems utility is superfluous, demand functions telling us all we need to know. A similar attitude was taken both by Barone (1908) and Moore (1914). In all his work on statistical demand curves, Moore found no help from utility theory.

The most significant exponent of this approach was, however, Samuelson (1938a) with his theory of revealed preference. Arguing that the hypothesis of a diminishing marginal rate of substitution introduced by Hicks and Allen was just as empty a concept as utility, Samuelson proposed to base consumer theory on a different set of assumptions, more directly related to observable behaviour. To do this he started from the same assumption as Cassel, Barone and Moore: that given the set of prices individuals choose to purchase certain commodities. Where Samuelson went further was in making the further assumption (the axiom of consistency) that when a consumer chooses a certain bundle of goods it can be

inferred that he or she prefers this bundle to all other bundles which could have been chosen instead. This enabled Samuelson to derive all Hicks and Allen's results without using even indifference curves. With Samuelson's work, the movement towards stripping consumer theory of all inessentials, reducing it to a theory of choice, could be said to have reached its conclusion.

14.3 ECONOMIC EQUILIBRIUM

1870–1914

There is a sense in which the concept of an economic equilibrium, in which the prices and quantities are determined by the balancing of various forces, permeates the whole of classical economics, from Smith to Ricardo, Mill and Cairnes. This is not, however, the whole story, for two reasons. (1) The classical economists never gave a significant role to demand in determining equilibrium prices.[18] (2) From its origins in Smith's *Wealth of Nations* classical economics was permeated by a concern with growth and development, the result being that the development of a theory of static equilibrium was a subsidiary theme. It was only in the period after 1870 that a thoroughly worked out system of statics did emerge.[19]

The clearest statement of the nature of economic equilibrium was undoubtedly that of Walras. Though important aspects of it were perceived by both Jevons and Menger, neither developed the concept so thoroughly: Jevons was more concerned with the principle of utility; Menger with explaining the essence of value. Walras, on the other hand, was concerend above all with the interdependence of various markets, a concern which led naturally into an analysis of general equilibrium. Though Walras, like the classical economists, analysed the evolution of an economy over time he made it clear that such analysis was dependent on a prior analysis of static equilibrium.

More clearly than either his predecessors or his contemporaries, Walras provided a theory of general competitive equilibrium: firms were price takers, and in equilibrium earned only the normal rate of return on capital. Equilibrium prices were determined by on the one hand consumers' maximization of utility, and on the other by the technical coefficients describing the ability of firms to transform inputs into outputs. Although neglected in England, Walras obtained influential followers in Europe, the most important being Pareto and Wicksell. It was Pareto who removed the theory's dependence on utility, arguing that the essence of the problem of economic equilibrium was "the opposition between men's tastes and the obstacles to satisfying them".[20] Wicksell, on the other hand, integrated Walras's theory of equilibrium with Böhm-Bawerk's theory of capital, extending it to provide a marginal productivity theory of distribution.

Very different was the Marshallian approach, not so much because of Marshall's preference for partial as opposed to general equilibrium analysis,

as because of his preference for a more realistic form of analysis. Being wary of excessive abstraction, Marshall's equilibrium was not the static, perfectly competitive equilibrium of Walras, Pareto and Wicksell: in it firms were continually evolving, old firms gradually being replaced by new; there were imperfections of competition; and some of the changes which occurred when firms moved along their supply curves were irreversible. Though not too much should be read into this, Marshall's use of the term "normal profit" rather than "zero profit" as his condition for market equilibrium is symptomatic of his approach. Furthermore, his preference for short chains of reasoning and his desire for realism worked against his analysing the logic of a general competitive equilibrium in the same way as his European contemporaries. His description of a full competitive equilibrium is brief, and in an appendix, not in the main text of the *Principles*.

Although changes were taking place even in Marshall's long period, there was in the background the notion of a full, stationary equilibrium, one in which events can be correctly anticipated an indefinite time beforehand. This was, however, a concept of which he was very critical, arguing

> it is to this cause more than any other that we must attribute that simplicity and sharpness of outline, from which the economic doctrines in favour in the first half of this century derive some of their seductive charms as well as most of whatever tendency they may have to lead to false practical conclusions.[21]

Despite being critical of the assumption of a stationary state Marshall did generalize the concept to cover balanced growth, a lead taken up by Cassel. On the whole, however, the problem of dynamics was set aside.

There was also a third approach to the question of competitive equilibrium, one which though not influential at the time, became important in the 1950s. This was Edgeworth's analysis of competition in terms of bargaining. Starting with a bargain between two individuals he derived the contract curve, showing that the competitive equilibrium was one point on this curve. By gradually increasing the number of individuals involved in the bargain Edgeworth was able to show that as the size of the economy increased so the contract curve shrank towards the competitive equilibrium. Competitive equilibrium could thus be interpreted as the only feasible outcome in a bargain between an infinitely large number of individuals.

Despite the tendency to assume some sort of competition, whether the Walrasian "free competition" or the Marshallian "economic freedom", problems of monopoly and oligopoly were not neglected. Cournot's influence was pervasive. His theory of monopoly was taken over by Marshall, who extended it by a more thorough analysis of the structure of costs, and a recognition of the fact that static optimization may be inadequate to describe a monopolist's behaviour. In Schumpeter's words, Marshall

> added little, if anything, to Cournot's analytic skeleton, but ... he developed from it ... an *economic analysis* that almost dwarfed both that skeleton and the technically superior performance of a later age.[22]

Cournot also influenced the period's discussion of oligopoly. His solution, involving the assumption that each of a pair of duopolists takes his rival's output as given, was criticized by many economists, in particular by Bertrand, Marshall and Edgeworth. Edgeworth's contribution is the most important here, for he showed that equilibrium would be indeterminate, a result which held not only when two firms produced the same good, but also when demands for their two, different, products were related.

These discussions of monopoly were never, however, integrated into the discussions of general economic equilibrium. Pareto rejected Walras's confining of pure economics to the study of perfect competition, aiming at greater generality. Despite this, however, he was unsuccessful in incorporating monopolistic elements into his general equilibrium analysis, this relating only to a competitive economy. Superficially, Marshall was more successful in dispensing with the over-simplification of perfect competition. This success was, however, to some extent more apparent than real, for it was only through his being imprecise about his exact assumptions that he was able to do this. The *formal skeleton* of his general equilibrium analysis was as reliant of the assumption of perfect competition as those of his contemporaries.

Chamberlin's monopolistic competition

In the 1920s and 1930s considerable progress was made in the analysis of economic equilibrium, the most notable developments being those sometimes described by the phrase "the monopolistic competition revolution". Two works were particularly influential: Chamberlin's *Theory of Monopolistic Competition* (1933) and Robinson's *Economics of Imperfect Competition* (1933a). Common to both theories, although derived independently, were the use of marginal revenue curves and the so called "tangency condition": if there is free entry to a market, but firms face downward-sloping demand curves for their own products, equilibrium will be attained where each firm's demand curve is tangent to its average cost curve. This implies that in equilibrium there will typically be excess capacity, with firms operating below the output at which average cost is minimized. But despite these similarities, and despite the concern of both Chamberlin and Robinson with the general theory of value, their approaches were fundamentally *very* different.

Chamberlin's central concern was to extend Marshall's work to deal with market structures characterized by advertising and product differentiation. For Chamberlin, firms controlled not only the price of their product, but also its quality and the amount of effort they put into advertising and selling it. Chamberlin was led, therefore, to classify markets not merely according to the number of sellers, but *also* according to the degree to which firms' products were differentiated.[23] Product differentiation implied that each firm, however great the number of its competitors, would have a certain

amount of monopoly power. In such a world of differentiated products the theory of monopoly, Chamberlin argued, will seem appropriate to explain the market for each firm's product. Competition, however, is not eliminated, for the behaviour of firms producing substitutes will affect demand for a monopolist's product.[24] Because each firm's product is different, there is no reason to expect that competition could ever eliminate monopoly altogether.

Several features of Chamberlin's approach are worth emphasizing. The first is the attention he paid to oligopoly. After arguing that competition and monopoly should not be separated, he turns, on the first page, to the theory of oligopoly.[25] In his discussion of oligopoly, the mutual dependence of oligopolists is emphasized. Monopolistic competition arises where the number of sellers is sufficiently large that such interdependence can be neglected. Secondly, the "tangency condition", to which so much attention has been paid, was, for Chamberlin, merely a special case. To obtain this special case it was necessary to make what he described as "certain heroic assumptions": in particular that each member of the group of firms being considered faced the same demand curve and the same cost conditions as other members of the group.[26] For Chamberlin, therefore, the tangency condition was used "mainly as an expositional device, of only limited direct applicability".[27] Thirdly, Chamberlin assumed that firms produce products which are genuinely different from the products of other firms. Thus monopolistic elements are not simply the result of irrational behaviour on the part of consumers. This meant that it was not possible to infer that monopolistic competition implied waste: perfect competition was not available in such a situation.

Chamberlin's work succeeded in showing that economists had to analyse a variety of market structures, not simply perfect competition and monopoly. It is on account of his destroying "the bold generalizations of Marshallian price theory" that Blaug concludes that "we are justified in speaking of a Chamberlinian Revolution in modern microeconomics in just the same way that we speak of a Keynesian Revolution in macroeconomics".[28] Chamberlin, however, never succeeded in providing a more general theory of value to supplant the competitive theory, for his theory remained at the level of partial equilibrium analysis. This point comes across most clearly if we consider Triffin's *Monopolistic Competition and General Equilibrium Theory* (1940) where he sees monopolistic competition theory as bridging the gap between the particular equilibrium analysis of Marshall, which focussed on the industry, and the Walrasian analysis, starting with the firm. Triffin claimed that monopolistic competition had abolished the "inner boundary" between the firm and the industry, going beyond Chamberlin in arguing that the concept of the industry, or "group" of firms, had to be dropped from value theory. There was, however, an inconsistency in this, for when, by abolishing the industry, the transition is made to a general equilibrium analysis, the macroeconomic implications of any changes have to be considered, something neither Chamberlin nor Triffin achieved.

Robinson's imperfect competition

Very different were the scope and purpose of Robinson's *Economics of Imperfect Competition*, as is evident from its opening words:

> "Among persons interested in economic analysis, there are tool-makers and tool-users." This book is presented to the analytical economist as a box of tools. It is an essay in the technique of economic analysis and can make only an indirect contribution to our knowledge of the actual world.[29]

Thus where Chamberlin had been concerned with realism, his geometry being in a sense incidental to his main inquiry, geometric techniques were prominent in Robinson's book. Robinson, however, did not confine herself to providing a box of tools: she drew conclusions about the real world. Where she did this her approach bore no resemblance to Chamberlin's. Product differentiation made no appearance in her book. Instead she used her techniques to draw conclusions about welfare. Thus when she came to consider monopsony in the labour market she admitted that despite her professed concern to be providing no more than a box of tools, "The temptation to stray from the path of analysis and to offer reflections of a moral character is here too strong to be resisted."[30] This shift of emphasis was complete in her final chapter in which, as she put it, "we are no longer occupied with the theory of Value, and have stepped over into the province of the Economics of Welfare".[31] She linked her work explicitly to Pigou's welfare economics.

In complete contrast to Chamberlin, Robinson emphasized the inefficiencies which resulted under imperfect competition, and on the exploitation of labour, this being defined as the difference between the real wage rate and the value of the marginal product of labour. Either monopoly in the product market, or monopsony in the labour market, would result in exploitation.[32]

The immediate context out of which Robinson's work emerged was discussions of Marshall's theory which took place in the late 1920s. These discussions were not concerned with the realism of Marshall's construct so much as with its internal logical consistency. The major contributors to the destruction of Marshallian theory which took place were Sraffa (1926), Robbins (1928) and Pigou (1928b).

Sraffa's argument was that in the long run, neither increasing nor decreasing returns to scale were compatible with partial equilibrium analysis of competitive industry. If there were economies of scale, these must be internal to the industry (otherwise they would not affect the supply curve) but external to the firm (otherwise a firm could expand indefinitely until it dominated the market). Since this category of economies was, according to Sraffa, hardly ever encountered, there was no way increasing returns could be compatible with competition. As for decreasing returns, these could arise, in the long run, only through increasing factor costs. But if factor costs rose when output increased, this would raise costs in other industries, and through affecting prices in other industries would affect demand. This, however, would violate the assumption, central to Mar-

shall's partial equilibrium analysis, that supply and demand were independent. So Sraffa concluded that returns to scale had to be constant, in which case a firm's output would be indeterminate. His solution to this problem was that economists needed to turn to the theory of monopoly.

Robbins' attack was on the concept of the representative firm. This was a device Marshall had used to deal with two types of problem: those arising from his supply curve not being a purely static construction (movements along the supply curve involved irreversible technical progress); and those arising from his assumption that each industry contained a variety of firms. The representative firm was a device used to ascertain how the industry would respond to a change in demand. It was not the average firm, but one judged to be typical in the relevant context. Robbins argued that not only was it difficult to recognize such a firm, for it was not an average which could be established statistically; but, more fundamentally, it contributed nothing to an understanding of economic equilibrium.

> Why then, when we come to deal with long period profit doctrine and the differences of managerial and business ability, should we find it necessary to consider a firm, an entrepreneur of average or typical efficiency? Just as units of a given supply may be produced on lands of varying efficiency, so their production may be supervised by businessmen of varying ability. What is normal profit for one will not be normal profit for another, that is all. *There is no more need for us to assume a representative firm or representative producer, than there is for us to assume a representative piece of land, a representative machine, or a representative worker.* All that is necessary for equilibrium to prevail is that each factor shall get as much in one line of production as it could get in any other: as much, of course, including all advantages and disadvantages of work, hiring or investment.[33]

The crucial aspect of this argument is that Robbins is interpreting Marshallian economics in the context of a theory of static equilibrium. Such a change was apparent also in Pigou (1928), where Marshall's representative firm was replaced with the very different one of the "equilibrium firm". The equilibrium firm, for Pigou, was a firm such that if it were in equilibrium, the industry too would be in equilibrium. Although Pigou put this forward as a different way of describing Marshall's representative firm, it was a different concept.[34] From Pigou's equilibrium firm it was but a short step to replace Marshall's notion of the industry, comprising a spectrum of different firms, with the simpler concept of an industry made up of a set of identical firms.

The final attack on Marshallian theory came from Harrod and Robinson. Harrod (1930) distinguished between the firm's demand curve and the market demand curve, something necessary under neither monopoly nor perfect competition, after which he derived the marginal revenue curve, relating it to price by the formula: $MR=P\ (1-1/e)$ where e is the elasticity of demand.[35] Harrod noted[36] that this implies the breakdown of supply and demand analysis, for if the firm produces an output such that marginal revenue equals marginal cost, supply will depend not simply on price but also on the elasticity of demand. This point was taken up by Robinson (1932, 1933a) who argued that there was no way of salvaging the notion of a

supply curve for an industry. The effect of a shift in demand on output would depend on how it affected the elasticity of demand for each firm in the industry.[37] The basis of Marshallian economics, namely the concept of a long period normal price dependent on supply and demand, was undermined.

Important though these results were, it can be argued that far more significant was Robinson's method. As were the articles of Robbins (1928) and Pigou (1928), *The Economics of Imperfect Competition* was an exercise in the theory of static equilibrium, premised on the assumption that "each individual acts in a sensible manner ... from the point of view of his economic interests".[38] Where Marshall had found a place for other motives, even altruism, Robinson postulated simple profit maximization:

> It is the assumption that any individual, in his economic life, will never undertake an action that adds more to his losses than to his gains, and will always undertake an action which adds more to his gains than to his losses, *which makes the analysis of value possible*. ... With bricks of this *one simple pattern* the whole structure of analysis is built up.[39]

This approach is closer to Austrian equilibrium analysis than to Marshall's "realistic" approach.

Robinson's simplifying assumptions simply cast aside many of the problems dear to Marshall, her analysis being conducted at a much higher level of abstraction. By way of explanation she pointed out that, though she did sometimes stray from this narrow path, her purpose was not to tackle real world problems, but to provide a "box of tools" for economists to use. In this she succeeded, her tools being the standard fare of contemporary microeconomics textbooks. The contrast between the old and the new methods is clearly revealed in an exchange between Shove and Robinson in the 1933 *Economic Journal*. Shove writes:

> So long as we are content with a rough and ready indication of the forces at work, we can keep fairly near to the facts: but any attempt to make our treatment exact is apt to lead either to a degree of abstraction which renders the analysis inapplicable to the actual phenomena we set out to explain or to a degree of complication which makes it cumbrous to use.[40]

Robinson's reply acknowledges the fundamental difference in approach:

> Indeed it is obvious that his realistic method of analysis and my highly formalised method do not operate in the same *terrain*, and any argument which turns upon the *results* obtained from such different sets of assumptions must in the nature of the case be idle[41]

Although this change in method had far reaching implications for the development of economic theory, it can be viewed in sharply contrasting ways. One interpretation[42] stresses the failure of the "new establishment" to deal with the vitally important problems of time, information and uncertainty with which Marshall had at least tried to grapple. The limitations of his approach can be put down to "methodological difficulties which could not be solved, only lived with".[43] Concern with an internally consistent

model of economic equilibrium is potentially dangerous, since the conditions under which it will be possible to make generalizations about the factors determining equilibrium will never be met. Economists, on the other hand, whose first priority is the internal consistency of a theory, even if this implies making assumptions which go against empirical observation, view the change in method rather differently. Samuelson, for example, though he criticizes those economists who made such a fuss about "discovering" marginal revenue in the 1930s, has written scathingly about Marshall's approach.[44]

Oligopoly

The attention which has been paid to imperfect competition should not be allowed to obscure the fact that in the 1920s and 1930s oligopoly received considerable attention. Chamberlin's interest in oligopoly has already been noted. The starting point for most discussions of oligopoly, as for that of Chamberlin, noted above, was the duopoly models of Cournot, Bertrand and Edgeworth.[45] A concept which helped to clarify the differences between various models was that of "conjectural variation": the amount by which a duopolist conjectures that his rival will alter her output should he increase his own output by one unit.[46] Cournot had assumed conjectural variations to be zero, but other assumptions can be made to give an immense variety of results. Of particular interest is Stackelberg's model (1933), in which one duopolist (the follower) has a conjectural variation of zero, whilst the other (the leader) takes as her conjectural variation the amount by which the follower will actually change his output.

The theory which gained the widest support was that of the kinked demand curve. This was implicit in Chamberlin's analysis, for he had distinguished two demand curves, one in which the industry's output was held constant when a firm changed its price, the other in which other firms varied their output. It was not far from these two, intersecting, demand curves to the kinked demand curve. The kinked demand curve was explicit in Robinson's *Economics of Imperfect Competition* (1933a).[47] The theory, however, received wide attention only with its appearance in 1939 in the work of Sweezy (in the US) and Hall and Hitch (in the UK),[48] even though their interpretation of it was no different from that of Robinson. In 1947 Stigler was able to claim that this theory had gained wide acceptance, some economists making it *the* theory of oligopoly price.[49]

The crucial assumption is that oligopolists expect competitors to match price cuts, but not price rises, which means that firms' demand curves will be less elastic for a fall in price than for a rise, the demand curve being kinked at the current price. Apart from its simplicity the theory had several features in its favour. (1) Through causing the marginal revenue curve to be discontinuous at the current price it explained why oligopolists' prices would be insensitive to changes in demand or costs. Sweezy went further, using the theory to explain why prices would be more likely to rise in response to an increase in demand, than to fall in response to a fall. (2) It was

possible to relate the theory to evidence which Hall and Hitch obtained from a survey of businessmen's pricing policies. This showed that businessmen neither understood the concept of the elasticity of demand, nor considered it relevant to their pricing decisions. On the other hand, there was a belief that price cuts, but not price rises, would be matched by competitors.

Perfect competition

In contrast with the theory of imperfect competition, the theory of competitive equilibrium was comparatively well worked out by 1914. Several important developments nonetheless occurred in the period up to 1939.

One development was work on the conditions required for the existence of a competitive equilibrium. The stimulus here came from Cassel's (1918a) reformulation of Walras's system of equations describing a general competitive equilibrium. It can be argued that Cassel's system of equations contributed little in itself, its significance resting solely in the research it stimulated. Three papers appearing in 1932–1933 showed that the problem of the existence of an equilibrium required more than the counting of equations.[50] Neisser (1932) showed that equilibrium prices might be negative; whilst Stackelberg (1933) showed that if there were fewer commodities than factors there might be no set of outputs such that all factors were fully employed. But arguably more important was the contribution of Zeuthen (1933) who pointed out the necessity of modifying the supply = demand equilibrium condition by allowing for the possibility that the supply of a commodity might exceed the demand in equilibrium, provided that the price of the commodity was zero. Once this modification was introduced Neisser's and Stackelberg's problems vanished. It was at the same time that the first steps were taken towards the more rigorous treatment of the problem of existence characteristic of modern theory. The need for this was perceived by Schlesinger (1933), the proofs of existence being supplied by a mathematician, Wald, who analysed the problem for a variety of models, including Cassel's model, and a model of pure exchange.

Although, however, the concept of competitive equilibrium was widely used, the notion of *perfect* competition arose only in the 1920s and 1930s, for it was only as economists tackled imperfect competition that perfect competition was properly defined.[51] It was Pigou (1928b) who first drew the now familiar U-shaped average cost curve with its associated marginal revenue curve, and who defined perfect competition in terms of equality of price with both marginal and average costs. The concept of perfect competition arose naturally out of the theory of imperfect competition: if an industry is made up of firms with identical cost conditions, and if the number of firms is sufficiently large for firms' demand curves to be completely elastic, the result is perfect competition. Perfect competition was thus a limiting case of imperfect competition. In Chamberlin's world where firms all produce different outputs, the concept of perfect competition arises less easily.

Although the concept of perfect competition, with its set of identical firms in equilibrium with horizontal demand curves tangential to identical U-shaped average cost curves, originated only with Chamberlin and Robinson, it can be argued that the essentials of this model were implicit, as was so much else, in the work of Cournot (1838). Of the economists considered in section II above, the one who was interested in the relation of firms to markets was Marshall. Marshall, however, chose to go in a different direction. His contemporaries, on the other hand, though they developed the theory of competitive equilibrium, were interested in issues other than that of the relationship of the firm to the industry. Thus, although it is correct to speak of a well-developed theory of competitive equilibrium before 1914, it is less correct to describe this as a theory of perfect competition in the modern sense.

Finally we come to Hicks's *Value and Capital* (1939a), a work which did much to revive interest in general equilibrium theory, not least because it explained general equilibrium theory in terms accessible to the non-mathematician. In this book Hicks concentrated on the case of perfect competition. His reasons for this were stated exceptionally clearly. After considering the equilibrium of a monopolist, he writes:

> So far, so good; yet it has to be recognised that a general abandonment of the assumption of perfect competition, a universal adoption of the assumption of monopoly, must have very destructive consequences for economic theory. Under monopoly the stability conditions become indeterminate; and the basis on which economic laws can be constructed is therefore shorn away. ... It is, I believe, only possible to save anything from this wreck – and it must be remembered that the threatened wreckage is the greater part of general equilibrium theory – if we can assume that the markets confronting most of the firms with which we shall be dealing do not differ very greatly from perfectly competitive markets. ... We must be aware, however, that we are taking a dangerous step. ... Personally, however, I doubt if most of the problems we shall have to exclude for this reason are capable of much useful analysis by the methods of economic theory.[52]

Hicks saw the main contribution of *Value and Capital* as lying in its attempt to combine the static analysis of Pareto with the dynamics of the Swedish economists, in particular Myrdal and Lindahl, whose influence on him was strong.[53] The first part of the book was concerned with presenting the theory of static equilibrium, reworking the Paretian theory of equilibrium in terms of the new consumer theory he had developed with Allen. Of particular importance was his analysis of stability, which related stability to the amount of complementarity in the economy. When it came to dynamics, Hicks adopted Lindahl's approach of studying a temporary equilibrium, one in which the stock of capital and the state of expectations were given. Any one period was linked to the past by the capital stock it inherited, and to the future by expectations. In this context Hicks was able to bring in problems of capital, interest and money, and was able to analyse the stability of the economic system as a whole.

Although Hicks saw the link between statics and dynamics as crucial, it was his interpretation of Keynes' *General Theory* as a miniature, aggregative general equilibrium model that attracted more attention. In his review of

the *General Theory* (1936) Hicks singled out Keynes' "method of expectations", comparing it with the method of Myrdal and Lindahl, whilst in his famous article "Mr. Keynes and the classics" (1937) he expounded the *General Theory* as a general equilibrium model with four markets: those for goods, money, bonds and labour. These ideas were taken further in *Value and Capital*, parts of which were strongly influenced by Keynes.

14.4 PRODUCTION AND DISTRIBUTION

Marginal productivity

It can legitimately be argued that marginal productivity theory as it is now understood dates from the last two decades of the nineteenth century. The idea of marginal productivity was not in itself new: the Ricardian theory of rent was a marginal productivity theory; and important aspects of marginal productivity theory can be found in Mill and, above all, Thünen. In addition, the work of both Jevons and Menger contains several important elements of marginal productivity. But several important developments occurred in the 1880s and 1890s, developments which served to transform marginal productivity theory into something qualitatively different from previous versions: (i) the concept of the production function emerged out of the classical laws of returns; (ii) marginal productivity became generalized to apply to all factors; and (iii) these developments permitted the production function to be set alongside the utility function, with factor prices being determined in substantially the same manner as prices of products. It was as a result of the extension and generalization of marginal ideas in this way that marginal analysis became generally accepted. Until the 1890s, despite the work of Jevons, Menger and Walras, classical influence was still strong.

The controversy over wages

"The great social question" concerning the relationship of capital and labour was a burning issue in the second half of the nineteenth century, in both England and the United States. The classical economists had used the wages fund theory to argue that wages were determined irrespective of the actions of either capitalists or trades unions, this providing the wage fund theory with a practical significance exceeding its significance in classical theory as a whole. It was because of this that Mill's "recantation" in 1869 caused such a stir. But the wage fund survived even Mill's recantation for not only did Cairnes, amongst others, continue to support it, but there was no agreed doctrine to replace it.

It was through the controversy which took place over the wages fund in the 1880s and 1890s that many of the most important contributions to marginal productivity theory emerged. The starting point for this debate was Walker's writings on the wages fund in the mid 1870s. Sidgwick, in a review (1879) of Walker's *The Wages Question* (1876) challenged Walker to

produce a positive theory to replace the wage fund theory, to which Walker responded in 1887, this setting off an extensive debate, in particular in the pages of the *Quarterly Journal of Economics*, a debate in which both American and British economists took part.[54] Many aspects of the debate are no longer interesting, but two aspects are, for they illustrate the changes which took place at this time in the theory of static equilibrium.

It came to be accepted that the time element was crucial, for defence of the wages fund involved assuming a lag in the production process analogous to the lag between input (sowing) and output (harvest) in an agricultural economy. A variety of economists defended the wages fund on the ground that this time dimension to the economic process could not be neglected. Indeed, only two economists denied such a lag and rejected the wages fund theory altogether: Clark and Marshall, and of these Marshall took a long time to emancipate himself from the concept.[55] Despite the enormous influence of Clark and Marshall, however, few economists in the 1880s and early 1890s went along with their rejection of the time element in production. Yet by the mid 1890s the wages fund was a dead issue. How did this come about? One explanation is the growth of marginal productivity theory, which enabled the problem of wage determination to be treated as a problem in statics. The other was Böhm-Bawerk's theory of capital, as a result of which the problem of time came to be seen as a problem in capital theory rather than the economics of wages. Discussions of time simply dropped out of the discussion of wages.

But perhaps the most important aspect of this controversy is that it is out of it that important contributions to marginal productivity theory emerged. Three papers are of particular importance. Wood (1888, 1889), in arguing that demand for labour was, like anything else, a function of price, worked out the essence of a marginal productivity theory of distribution, but was largely neglected.[56] And in 1891 Hobson and Clark, as the titles of their articles, "The law of the three rents" and "Distribution as determined by a law of rent", suggest, explicitly generalized the theory of rent to apply to all factors, not simply land.

Although outside the mainstream of academic discussion, there is a side to the discussion of distribution that cannot be neglected – Henry George's *Progress and Poverty* (1879). George was concerned to argue that the problem of rent, and the unearned increment of land value, was the principal social problem of the time. To deal with it he advocated the use of a single tax – a tax on rents. Although he contributed little, if anything, to economic analysis,[57] his ideas stirred up a ferment of popular discussion, not only in the US, his home, but also in Britain.[58] It is partly due to George's influence that discussions of rent were so widespread at this time.

Marginal productivity and the production function

Though English economists were active and important participants in the debates over the wages fund this was primarily an American controversy, in that not only was it conducted primarily in American journals, but also in

that most of the participants were American. Overlapping these discussions, however, was another debate, involving primarily European economists, one which examined marginal productivity theory in rather a different way.

In the same way that the first American statement of marginal productivity theory, that of Wood, was neglected, so too were the earliest European statements: those of Berry (1891) and Edgeworth (1889). Berry's contribution is worth summarizing in some detail. He writes down a production function, $f(g_1, g_2,\ldots l_1, l_2,\ldots c)$ giving a firm's output as a function of the inputs of different types of land (g_k), different types of labour (l_j) and capital. This is then used to derive marginal productivity equations for each factor:

$$\begin{aligned} p_i df_i/dg_k &= r_k \\ p_i df_i/dl_j &= w_j \\ p_i df_i/dc &= i \end{aligned}$$

p_i, r_k, w_j, and i being the prices of output, rents of land, wage rates and the interest rate. These equations are then supplemented by demand equations for goods, supply equations for factors and the condition that all factors are fully used. The result of this was a general equilibrium system with as many equations as unknowns. Capital, labour and land were treated symmetrically, leading to the following, anti-Ricardian conclusion:

> There is no more justification for assuming wages to be measured by the produce of the labourer working on the margin *without capital* than for assuming interest to be measured by an amount of capital *without labour*.[59]

In view of subsequent discussion it is worth noting Berry's method of ensuring exhaustion of the product: any difference between the value of the output and the value of the inputs constituted a surplus, appropriated by the entrepreneur. Competition would equalize the surplus received by entrepreneurs of equal ability.

Two features of Berry's work, his symmetric treatment of factors, and his treatment of marginal productivity as part of the general theory of value, are even more explicit in the book which aroused the greatest controversy in the 1890s: Wicksteed's *Essay on the Co-ordination of the laws of Distribution* (1984). After a preface in which he extolled the virtues of a mathematical approach making the production function explicit, he states his case very clearly:

> In investigating the laws of distribution it has been usual to take each of the great factors of production such as Land, Capital and Labour, severally, to enquire into the special circumstances under which that factor co-operates in production, the special considerations which act upon the persons that have control of it, and from all these considerations to deduce a special law regulating the share of the product that will fall in distribution to that particular factor.

> Now as long as this method is pursued it seems impossible to co-ordinate the laws of distribution and ascertain whether or not the shares which the theory assigns to the several factors cover the product and are covered by it. ... As long as the law of rent, for example, is based on the objective standard of the fertility of land, while the law of interest is based on the subjective standard of estimate of the future as compared with the present, it is difficult even to conceive any calculus by which the share of land and the share of capital could be added together and an investigation then instituted as to whether the residual share will coincide with what the theory assigns as the share of wages. But it is obvious that such a co-ordination must be within the purview of economic theory.[60]

As for the method used, he points out that "the modern investigations into the theory of value" provide this, "the law of exchange value [being itself] the law of distribution of the general resources of society".[61]

Where Wicksteed went beyond other authors, and where controversy arose, was in his attempt to show that factor shares determined by marginal productivity would completely and exactly exhaust the product if the production function exhibited constant returns to scale, or linear homogeneity.[62] His proof of this was rather clumsy, but was improved upon by a reviewer, Flux (1894) who, for the first time, used Euler's theorem to analyse the problem, an approach which has now become standard.

Reaction to Wicksteed's theory was mixed. Amongst continental economists Barone, Pareto and Walras were very hostile, attacking the assumptions of linear homogeneity, and full substitutability of factors of production. These criticisms were echoed by Edgeworth. Many of their criticisms were misconceived, but some progress was made. In particular Barone derived a product exhaustion theorem from the assumptions that (i) firms minimize cost, and (ii) cost and selling price are equal.[63] Marshall too was critical, but in a more guarded manner. He had been using marginal productivity ideas since his *Economics of Industry* in 1879, but he warned against thinking that marginal productivity could provide more than a part of a theory of wages because it analysed only factor demand. Despite this apparent hostility, however, successive editions of the *Principles* conceded more and more to the marginal productivity theory. For example, Wicksteed's theory about product exhaustion was incorporated into the third edition. The main contribution towards developing the marginal productivity theory came, however, not from Wicksteed's critics, but from his most consistent supporter, Wicksell.[64] Wicksell approached the problem by asking whether it mattered whether the landowner was the entrepreneur, hiring the workers, or whether the workers were the entrepreneurs, renting the land. But his main contributions were twofold: (i) he saw increasing, decreasing and constant returns not as alternatives but as applying to different regions of the production function. Thus Wicksteed's theory applied not because returns were constant throughout, but because firms moved to the portion of the production function exhibiting constant returns to scale. (ii) He argued that product exhaustion was an equilibrium

condition rather than a condition which would be satisfied all the time. This interpretation was later endorsed by Hicks in his *Theory of Wages* (1932).

It was out of these discussions that the concept of the production function emerged, its derivatives describing the classical laws of returns. Nowadays these ideas may seem obvious, but this was not the case at the time, as is evidenced by the confusions abounding in many of the contributions made by otherwise distinguished economists. One reason, perhaps, for the controversy was Wicksteed's attempt to explain too much in terms of a single property of the production function. This accounts for Edgeworth's comment,

> There is a magnificence in this generalization which recalls the youth of philosophy. Justice is a perfect cube, said the ancient sage; and rational conduct is a homogeneous function, adds the modern *savant*.[6]

Wicksteed did not have a proper theory of the firm and, as Wicksell later showed, an understanding of the theory of the firm was crucial to an understanding of Wicksteed's product exhaustion theorem.

So far the term production function has been used as though it referred to a single concept, but this is not the case. A production function can, at one extreme, refer to the function relating a single firm's output to a detailed list of all the inputs it uses. At the other extreme the concept relates to a whole economy, relating total social production to broad aggregates of land, labour and capital. The aggregation involved in the latter, both across firms and across factors, makes it a very different, and much more problematic, concept than the non-aggregative micro production function. We have seen Berry's use of the production function to describe the firm's production possibilities, a usage shared by Marshall and Walras, the only major economists of the period, according to Schumpeter, who were entirely free of the notion of an aggregate production function. Of the other authors discussed above, Clark, Wicksteed, and Wicksell argued explicitly in terms of an aggregate production function.

The classic statements of a marginal productivity theory of distribution based explicitly on an aggregative social production function came in the 1930s: Hicks' *Theory of Wages* (1932) and Douglas's *Theory of Wages* (1934). Using an aggregate production function Hicks analysed distributive shares in terms of the newly-developed concept of the elasticity of substitution. The elasticity or substitution, which measures the curvature of an isoquant at any point, determines whether an increase in, for example, labour will increase or reduce the share of labour in national income. As was also the case with Douglas's theory, this theory provided a very clear account of a static, aggregative marginal productivity theory of distribution. Douglas, along with his collaborator, Cobb, used the specific form of production function that has come to be named after them: $Y = AK^{\alpha}L^{1-\alpha}$.[66] Because the Cobb–Douglas production function has an elasticity of substitution equal to one, factor shares will be determined solely by the parameter α of the production function. This function formed the basis for Douglas's extensive econometric research on the distribution of income.

14.5 CAPITAL AND INTEREST

In the theories of distribution discussed above the share of capital is determined by its marginal product, but this makes no sense until we have defined what we mean by capital, and how it enters the production function. This was the subject of enormous controversy, especially in the closing decades of the nineteenth century when Böhm-Bawerk and Clark advocated diametrically opposed conceptions of capital, and in the 1930s when the main protagonists were Hayek and Knight. In both periods controversy centred on whether production should be viewed as a process in which inputs were applied at different dates, capital being measured by the average period for which resources were invested in the production process; or whether production should be viewed as a synchronous process, the current stock of capital and the current stock of labour determining the current level of output. Böhm-Bawerk and Wicksell were the outstanding advocates of the former view; Clark of the latter.

Böhm-Bawerk and Wicksell

The idea of looking at production as a process involving time has a long history. We find in Ricardo the idea that the value of a commodity is made up of the value of labour input (this comprising both labour used directly and labour used indirectly, the latter being "stored up" in capital), plus an interest charge proportional to the time between the application of the inputs and the production of the output. Indeed, this, together with the fact that production processes for different goods were of different lengths, was one of the reasons why the labour theory of value would not work. It was a view of capital even more clearly specified by Jevons, who went so far as to write down a function relating output to the average period of production (the average time for which labour is invested), obtaining the rate of interest as the derivative of this function: the rate of interest was the marginal product of a lengthening of the period of production. The economist who, above all others, is associated with this theory is, however, Böhm-Bawerk, who advocated it in two widely read books: *History and Critique of Interest Theories* (1884) and, above all, *Positive Theory of Capital* (1889), both translated into English soon after publication (1890 and 1891 respectively).

There were two reasons, apart from the nature of the theory itself, why Böhm-Bawerk's work was so extensively criticized. One was that the quality of his insights far exceeded the quality of his technique. He made technical blunders onto which his critics could fasten. Schumpeter claimed that when all the necessary corrections were introduced into Böhm-Bawerk's theory nothing was left of it except the essential idea.[67] The other reason for criticism of Böhm-Bawerk was that he mixed together two rather different objectives: to provide a causal explanation of interest; and to provide a model in which, given the supply of labour and the stock of the means of subsistence, the rate of interest, wage rates and the period of production were simultaneously determined. These two parts of his theory

will be considered in turn, but before that we need to examine his concept of the period of production.

Though Böhm-Bawerk held to a physical concept of capital (the stock of intermediate, including subsistence, goods) he did not regard capital as a third factor to be placed alongside labour and land, the two "original" factors, the stocks of which were determined by non-economic factors. Capital was, for Böhm-Bawerk, productive because it permitted the adoption of more roundabout methods of production, which increased the output from the "original" factors. He assumed that the adoption of more roundabout methods involved an increase in the average time for which capital was invested. Because it leads to more roundabout methods being used, an increase in the period of production, which Böhm-Bawerk used as his measure of the capital stock, would raise the productivity of the two original factors, labour and land, but it would do so at a diminishing rate. The function of capital was to increase the productivity of the other fasctors of production. Competition would ensure that these received not their marginal products, but the discounted values of their marginal products, the remainder constituting interest.

This raises the question of why, if capital is productive, and if its supply is not fixed by non-economic factors, the period of production should not be extended indefinitely, to the point where interest disappeared. Böhm-Bawerk adduced three reasons – his "reasons for interest":

(i) different circumstances of want and provision in the present and the future (if people expect to be better off in the future they will value present consumption more highly than future consumption);
(ii) undervaluation of the future (myopia, limited will power or the uncertainty of life);
(iii) the technical superiority of present over future goods (the greater productivity of longer production processes).

Though Böhm-Bawerk claimed that these were three independent reasons for interest, it can be argued, as did Fisher,[68] that the third is not: the rate of interest depends on supply and demand for loans, but Böhm-Bawerk's third reason affects only demand. Together, however, the three reasons do provide an explanation of interest.

Finally we come to Böhm-Bawerk's explanation of how the interest rate is determined. His starting point is the assertion that "The exchange of present goods for future goods, which constitutes the source of the phenomena of interest, is merely one special case under the rubric of exchange of goods in general."[69] So he starts, following Menger, with isolated exchange, where the price of a loan will be set somewhere between its value to each of the two individuals involved. But of more interest is his explanation of interest under market conditions. The main points here are illustrated by his simplest case, which is worth examining in some detail. He makes the assumption that capitalists are the sole suppliers and demanders of funds, and that workers are the only consumers. Thus the sole function of the subsistence fund is to provide for wages. Rather than use his numerical examples, however, we will use Figure 14.1 which

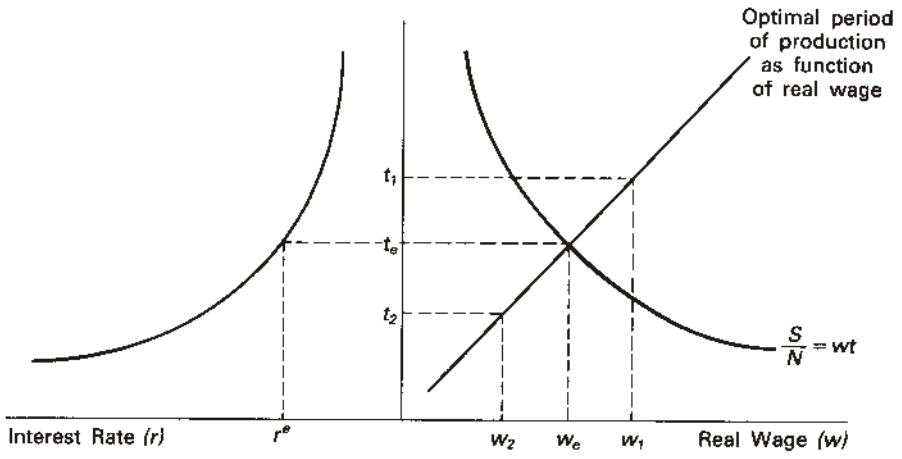

FIGURE 14.1 *Böhm-Bawerk's Theory of Interest*

represents graphically the relationships Böhm-Bawerk assumed. Three curves need to be considered. (i) As the real wage increases, so the optimal (profit maximizing) period of production will increase: at higher real wage rates firms will wish to use more capital-intensive methods. (ii) Corresponding to this is an inverse relationship between the interest rate and the period of production: as the production becomes more capital-intensive, so the marginal product of capital falls. (iii) Finally there is the condition that the stock of subsistence goods should equal the number of workers multiplied by the real wage rate times the period of production. Suppose the real wage were too high, say w_1.[70] The optimal period of production will rise to t_1, implying a low rate of interest (not marked). Given this long period of production, however, a stock of subsistence amounting to $w_1 t_1 N$ would be required. This is greater than S (i.e. the point t_1, w_1 is to the right of $S/N=wt$). This means that capitalists will be unable to employ the whole labour force, and so the real wage rate will be bid down. Given the fixed subsistence fund employment will then be less than the supply of labour, so the real wage will be bid down. Similarly if the real wage were too low, say w_2, only a part ($w_2 t_2 N$) of the subsistence fund would be employed, so the real wage would be bid up. Three factors, therefore, determine the rate of interest:

1) The magnitude of the subsistence fund [S];
2) The number of workers the fund must support [N];
3) The gradation in the scale of increasing productivity that accompanies prolongation of the production period [the shape of the curves relating r, w and t][71]

The staunchest defender of this theory was Wicksell, who did more than clear up technical blemishes and provide a more elegant exposition, true as both of these are. His main achievement was to integrate Böhm-Bawerk's

theory with the Walrasian theory of general equilibrium, and to incorporate land as well as labour into the theory of capital and interest. But even Wicksell did not manage to clear up all the technical problems involved with this view of capital. Indeed, through his discovery of what is now called the "Wicksell effect" he opened up a path for later generations of critics of Böhm-Bawerk's and other theories of capital. Wicksell used this effect, whereby a change in the supply of capital causes a revaluation of the capital stock, altering the units in which it is measured, to explain why the marginal product of capital would usually be less than the rate of interest. The full significance of this, however, was not appreciated until the 1950s. Wicksell, for example, despite his good formal theory, ended up assuming that the marginal product of capital was equal to the rate of interest.[72]

Clark

Fundamental to Clark's criticism of Böhm-Bawerk, as well as to his own view of capital, is the distinction between capital as a fund of value, and capital goods. Capital is, according to Clark, "a permanent fund of productive wealth, expressible in money" but embodied in capital goods. In a stationary economy there will be a constant fund of capital, but the concrete capital goods in which this is embodied are continually changing: capital goods are wearing out and being replaced by others. He uses the analogy of a waterfall:

> A water-fall consists in particles of water. Can one say the same things of the fall that he does of the water? The water moves; the fall stays where it is. The water appears in globules condensed in the atmosphere, and it ultimately merges itself in the sea. The fall does not appear nor disappear. Capital goods are, like particles of water, vanishing elements. True capital is like the fall; it is an abiding element, owing its continuance to the constant wasting and replenishing of its substance.[73]

Whereas Böhm-Bawerk looked at individual capital goods, each of which has a period of production, Clark argued that the capital fund was far more important, having no periods but acting incessantly. Furthermore, where Jevons and Böhm-Bawerk saw the essence of capital as permitting production to be spread out over time, Clark argued the reverse: capital permits different stages of the productive process to be carried on simultaneously.

Though Clark and Böhm-Bawerk were advocating radically different views of capital they had one thing in common: both tried to measure capital by a single number, whether the period of production or the value of the capital fund. Both used highly aggregative models of production. It can be shown, as was done in the 1960s, that similar criticisms are applicable to both theories: in circumstances where the period of production fails, so too does Clark's concept of the capital stock.

Cassel

A third approach to the theory of capital was the one adopted by Marshall, and later in *The Nature and Necessity of Interest* (1901) by Cassel, namely to

define capital in such a way that it could be regarded as an original factor of production, standing alongside land and labour. The independent, or primary factor to which capital corresponded was found in "abstinence" (Marshall) or "waiting" (Cassel). For all his stress on time preference this was something Böhm-Bawerk had refused to do, seeing land and labour as the only two primary factors. Cassel's argument was that waiting had to be regarded as an independent factor: it could not be reduced to more elementary factors; and it could be substituted for other factors.[74] The interest rate was a price, determined by same factors as the price of any other good, arising because waiting was scarce, while its outcome (saving) was in demand for capital investment. To determine the rate of interest it was necessary simply to look at the demand and supply of waiting.

Fisher

Fisher too acknowledged a debt to Böhm-Bawerk, for while he objected to the latter's assumption that longer processes were more productive than shorter ones, he accepted most of the rest of his theory. Fisher's solution to the problem of interest, taking into account Böhm-Bawerk's discussion of time and the provision for wants at different times in the future, was

> that the rate of interest depends on the character of the income stream, – its size, composition, probability, and above all its distribution in time. It might be called a theory of *prospective provision of income*.[75]

This theory was expounded in *The Rate of Interest* (1907), later revised as *The Theory of Interest* (1930). His earlier work is also important, however. In *The Nature of Capital and Income* (1906) he defined capital value as the discounted present value of a future income stream. The rate of interest was the price linking the flow of income with the stock of capital value. And in *Appreciation and Interest* (1896) he investigated the distinction between real and nominal interest rates.

Fisher developed his theory in three stages. In the first he assumed that the income stream facing each individual was given. Given the rate of interest the consumer has to choose a consumption stream with the same present value as the given income stream, a point Fisher illustrated with the now-familiar diagram shown as Figure 14.2 (at this stage consider only the budget line and the indifference curves). The difference between income and consumption is the individual's demand or supply of loans, this depending on the interest rate. Market equilibrium requires that the interest rate be such as to equate demand and supply of loans in the market as a whole.

Borrowing and lending as in the example above are not the only ways of altering the income stream: the owner of "capital-wealth" has alternative uses to which he may put it. In Clark's terms, the capital fund may correspond to different physical capital goods, each of which yields a different income stream. Fisher, however, recognized that, "when an income stream is modified by a change in the use of the capital yielding it, its present value may not remain the same.".[76] For Fisher, unlike Clark, the

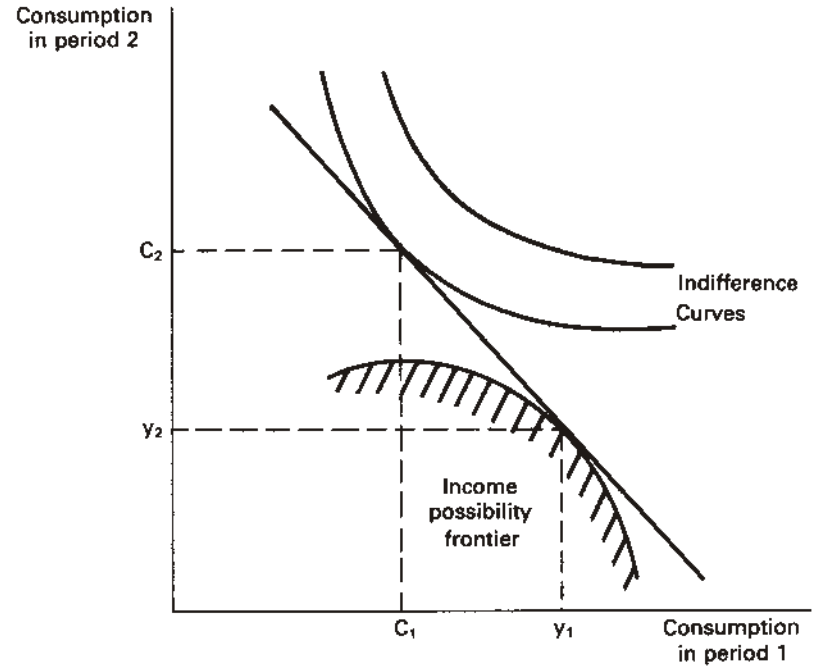

FIGURE 14.2 *Fisher's Theory of Interest*

value of the capital stock could not be regarded as a given factor of production. Given the ability to change the income stream through changing the use of the capital stock, which Fisher illustrated with the income possibility frontier shown in Figure 14.2, and the rate of interest, an individual's optimum will be that shown in Figure 14.2. As in the previous case demands and supplies of loans can be calculated for all the individuals in the economy, the equilibrium interest rate occurring where these sum to zero.

Finally there is the third stage where uncertainty is introduced. Fisher argues that the market will become segmented, different interest rates applying to different degrees of security and to different time periods. In addition, risk will increase time preference. But despite this Fisher failed to provide a detailed analysis to match his discussion of the determination of interest under certainty. He contented himself with arguing that risk introduced disturbing influences into the static scheme: for example, not all borrowers and lenders would face the same interest rate, and so their rates of time preference would differ. His conclusion, in *The Theory of Interest*, is worth quoting at length, for it indicates his view of what economic theory is capable of doing, and his view as to the limitations of economic analysis.

We must, therefore, give up as a bad job any attempt to formulate completely the influences which really determine the rate of interest. ... In short, the theory of interest in this book merely covers the simple rational part of the causes actually in operation. The other or disturbing causes are those incapable of being so simply and rationally formulated.[77]

Study of these disturbing causes had to be empirical and statistical, not rational and theoretical.

In conclusion, two further things are worth noting about Fisher's theory. In *The Theory of Interest* the analysis of production was supplemented by the important concept of the rate of return over cost, substantially identical to Keynes' marginal efficiency of capital. This is a concept founded on the notion, central to Fisher's work, of capital as the present value of an income stream. It was not a physical rate of return on capital, but a variable determined by comparing two streams of income. The second point is that, in contrast to the theories of both Clark and Böhm-Bawerk, Fisher's theory is microeconomic. Though, like Clark, he views capital as a fund of value, the similarity ends there: there is nothing of Clark's social capital in Fisher's work; nor anything paralleling Böhm-Bawerk's period of production.

Schumpeter and Knight

The only aspect of profits considered so far has been interest. Pure profits do not appear in static equilibrium because entrepreneurial activity reduces them to zero. The fact that entrepreneurial profits are zero in static equilibrium, however, does not mean that they are unimportant. J. B. Clark, for example, had argued that whilst entrepreneurial profits were a dynamic phenomenon, occurring only in disequilibrium, they were important because it was the prospect of such profits which provided the incentive to innovate. The economist who placed the greatest stress on this, however, was Schumpeter. Schumpeter, in his *Theory of Economic Development* (1912) argued that without innovations and technical change an economy would eventually settle down to a stationary state in which there was no uncertainty about the future. It is innovations which disturb this situation, these leading to unforeseen opportunities for profit.[78]

A similar view was later taken up by Knight in *Risk, Uncertainty and Profit* (1921), who used uncertainty, in a carefully defined sense, to explain not interest but pure profit, the surplus above the costs of all factors of production. He distinguished two types of uncertainty: a *measurable* uncertainty (such as in games of chance) which he called risk, and an *unmeasurable* uncertainty which, he argued, constituted true uncertainty. True uncertainty, when people do not know what they are doing, is the essence of profits, but as people learn, this uncertainty will be reduced, and with it profits.

Hayek and Knight

Debate over the theory of capital re-emerged in the 1930s, the stimulus being Hayek's *Prices and Production* (1931). In this book Hayek used the

Böhm-Bawerkian theory of capital as part of an explanation of the trade cycle. The characteristic of a cyclical upswing was seen to be a lengthening of the period of production, caused by a primarily monetary expansion. When monetary expansion ceased, according to Hayek, a crisis would emerge because long, capital-intensive production processes would become unprofitable and would be stopped. The result would be over-production of investment goods relative to consumption goods, unemployment emerging until labour could be redeployed.[79] Though it can be argued that the crucial factor here is the inflexibility of the capital stock rather than the Böhm-Bawerkian theory *per se*, this book started a debate on the capital theory used by Hayek, his most notable critic being Knight, whose first contribution appeared in 1933.[80]

Knight's criticisms of the Böhm-Bawerk/Wicksell/Hayek view were varied. He argued (i) that Böhm-Bawerk's distinction, taken from the classics, between primary and secondary factors was misconceived, the relation between capital and labour being "strictly mutual, co-ordinate and simultaneous"; and (ii) that there is no period of production with any determinate length or meaning. The justification for the former assertion is contained in the following quotation.

> In the historical view the creation of the productive system itself, including labourers as well as capital instruments, which in turn include "land", has been a cumulative, uninterrupted process of the hen-and-egg sort, going back as far as we care to trace it; in this process, moreover, all productive instruments existing at any time, including labourers, have participated on a joint co-operative basis.[81]

As for the second point, he had a variety of explanations. As indicated above, he saw production as a process with its origins infinitely far back, and with implications stretching infinitely far into the future. Associated with this was the impossibility of distinguishing between maintenance and new investment. Only if such a distinction is made can new processes being started be distinguished from old ones being continued. But more important than this was the claim that there was no necessary connection between roundaboutness and the stock of capital. Roundaboutness might increase without any increase in the stock of capital. As regards the stationary state, Knight accepted Clark's view of production as a process in which "the productive equipment of society yields want-satisfying services which are consumed as soon as they are created".[82]

Some of Knight's criticisms, such as his argument that a single infinitely-lived item of capital equipment makes the average period of production infinite, were unfounded. He was right, however, in claiming that the period of production could not, except under very special circumstances, be used to measure the capital stock.[83] As for his own view of capital, this was similar to Fisher's. Capital was regarded as identical to wealth, the capital value of a consumption stream. In a stationary economy this will equal the cost of the capital goods. It does indeed take time to change the capital stock, but the rate at which this can be done depends not on the structure of investment, the Böhm-Bawerkian answer, but on the amount of saving.

14.6 CONCLUSIONS

Although it remained firmly based on the ideas of Jevons, Menger, Walras, Marshall and Clark, the theory of economic equilibrium was by 1939 very different from that of the pioneers of marginal analysis. (1) By 1939 it had been made abundantly clear that the maximizing model of consumer behaviour did not depend on any hedonistic psychology. (2) Market equilibrium, both competitive and non-competitive, was now understood much better. (3) The application of mathematical techniques, which in the post-war period was to be used to produce even more fundamental changes in economic theory, had become firmly established. (4) There had been significant progress, notably by the Swedish school, and by Hicks, in analysing an economy characterised by time and uncertainty. (5) Finally, underlying all these developments were important developments as regards method.[84] Most economists had turned away from Marshall's attempt to blend history and economic theory, and from Menger's search for causal laws. When we take account of the increased interest in econometrics, the method of the new establishment in value theory was probably closer to that of Jevons than to that of any of his contemporaries. Taken together, these developments implied a theory of value radically different from that prevailing half a century before.