1 The neoclassical maximization hypothesis

> At present the maximization postulate has an unusually strong hold on the mind set of economists... Suffice it to say that in my view the belief in favor of maximization does not depend on strong evidence that people are in fact maximizers... The main argument against the maximization postulate is an empirical one – namely, people frequently do not maximize. Of course, this standpoint argues that while postulates simplify reality, we are not free to choose counterfactual postulates. Hence, from this point of view a superior postulate would be one under which maximizing behavior is a special case, but non-maximization is accommodated for as a frequent mode of behavior.
>
> Harvey Leibenstein [1979, pp. 493–4]

> If by rational we mean demonstrably optimal, it follows that conduct in order to be rational must be relevantly fully informed.
>
> George Shackle [1972, p. 125]

> The assumption of maximization may also place a heavy (often unbearable) computational burden on the decision maker.
>
> Herbert Simon [1987, p. 267]

The assumption of maximization is a salient feature of every neoclassical explanation. Obviously, then, if one wanted to criticize neoclassical economics it would seem that the most direct way would be to criticize the assumption of universal maximization. Several approaches have been taken. Harvey Leibenstein [1979] offered an external criticism. He argued for a ‘micro-micro theory’ on the grounds that profit maximization is not necessarily the objective of the actual decision-makers in a firm and that a complete explanation would require an explanation of intrafirm behaviour. He also gave arguments for why maximization of anything may not be realistic or is at best a special case. Similarly, Herbert Simon has argued that individuals do not actually maximize anything – they ‘satisfice’ – and

yet they still make decisions.[1] And of course, George Shackle has for many years argued that maximization is not even possible.

Some anti-neoclassical economists are very encouraged by these arguments, but I think these arguments are unsuccessful. For anyone opposed to neoclassical theory, a misdirected criticism, which by its failure only adds apparent credibility to neoclassical theory, will be worse than the absence of criticism. The purpose of this chapter is to explain why, although the neoclassical hypothesis is *not a tautology* and thus may be false, no criticism of that hypothesis will ever be successful. My arguments will be based first on the possible types of theoretical criticism and the logic of those criticisms, and second on the methodological status of the maximization hypothesis in neoclassical explanations.

TYPES OF CRITICISM AND THE MAXIMIZATION HYPOTHESIS

There are only two types of *direct* criticism of any behavioural hypothesis once its logical validity has been established. One can argue against the *possibility* of the hypothesized behaviour or one can argue against the *empirical* truth of the premise of the hypothesis. In the case of the neoclassical maximization hypothesis, virtually everyone accepts the logical validity of the hypothesis. For example, everyone can accept that *if* the consumer is a utility maximizer, then for the particular bundle of goods chosen: (a) the marginal utility is zero, and (b) the slope of the marginal utility curve at the point representing the chosen bundle is non-positive and usually negative.[2] That is to say, necessarily the marginal increment to the objective must be zero and falling (or not rising) whenever (i.e. without exception) the maximization premise is actually true. Of course, one could substitute the word 'profit' for the word 'utility' and the logic of the hypothesis still holds. With either form, (a) and (b) are the 'necessary conditions' for maximization. Note again that there are no 'sufficient conditions' for maximization. Rather, the maximization premise is the sufficient condition for (a) and (b).

Parenthetically, I should note that economists often refer to the *conjunction* of (a) and (b) as a sufficient condition for maximization. This is a common error.[3] Even if (a) and (b) are both true, only *local* maximization is assured. However, maximization in general (i.e. global) is what the premise explicitly asserts and that is not assured by (a) and (b) alone. I will return to this below when I discuss the methodological uses of the maximization hypothesis.

THE LOGICAL BASIS FOR CRITICISM

As stated above, there are two types of direct criticism of the maximization hypothesis: the possibilities criticism and the empirical criticism. In this section I will examine the logical bases of these critiques, namely of the possibilities argument which concerns only the necessary conditions and of the empirical argument which concerns only the statements which form the sufficient conditions. In each case I will also discuss the possible logical defense for these criticisms.

The possibilities critique: can the necessary conditions be fulfilled?

The possibilities critique builds on the difference between necessary and sufficient conditions. Specifically, what is criticized is the possibility of fulfilling all of the necessary conditions for maximization. Of course, this type of critique begs the question as to what are all the necessary conditions. Are there more conditions than the (a) and (b) listed above? Shackle, following Friedrich Hayek and John Maynard Keynes, argues that maximization also presumes that the knowledge necessary for the process of choosing the 'best' alternative has been acquired.[4] For Shackle, maximization is always a deliberate act. Shackle argues that for maximization to be a behavioural hypothesis (i.e. about the behaviour of decision-makers), the actor must have acquired all of the information necessary to determine or calculate which alternative maximizes utility (or profit, etc.) and he argues that such an acquisition is impossible, hence deliberate maximization is an impossible act.

Although this argument appears to be quite strong, it is rather elementary. A closer examination will show it to be overly optimistic because it is epistemologically presumptive. One needs to ask: Why is the possession of the necessary knowledge impossible? This question clearly involves one's epistemology – that is, one's theory of knowledge. The answer, I think, is quite simple. Shackle's argument (also Hayek's and Keynes') presumes that the truth of one's knowledge requires an inductive proof. And as everyone surely knows today, there is no way to prove one's knowledge inductively whenever the amount of information is finite or it is otherwise incomplete (e.g. information about the future).[5]

The strength of Shackle's argument is actually rather vulnerable. Inductive proofs (and hence inductive logic) are not necessary for true knowledge. One's knowledge (i.e. one's theory) can be true even though one does not know it to be true – that is, even if one does not have proof. But I think there is an even stronger objection to the 'true knowledge is necessary for maximization' argument. True knowledge is not necessary

for maximization! Consumers, for example, only have to think that their theory of what is the shape of their utility function is true. Once a consumer picks the 'best' option there is no reason to deviate or engage in 'disequilibrium behaviour' unless he or she is prone to testing his or her own theories.[6]

In summary, Shackle's inductivist argument against the possibility of a true maximization hypothesis is a failure. Inductive proofs are not necessary for true knowledge and true knowledge (by any means) is not necessary for successful or determinate decision-making. Maximizing behaviour cannot be ruled out as a logical impossibility.

The empirical critiques: are the sufficient premises true?

Simon and Leibenstein argue against the maximization hypothesis in a more straightforward way. While accepting the logical validity of the hypothesis, they simply deny the truth of the premise of the hypothesis. They would allow that if the consumer is actually a maximizer, the hypothesis would be a true explanation of the consumer's behaviour but they say the premise is false; consumers are not necessarily maximizers hence their behaviour (e.g. their demand) would not necessarily be determinable on that basis. Leibenstein may allow that the consumer's behaviour can be determined, but it is an open question as to what is the determining factor – utility, prestige, social convention, etc.? Simon seems to reject as well the necessity of determinate explanation although he does discuss alternative decision rules to substitute for the maximization rule.[7]

A denial of the maximization hypothesis on empirical grounds raises the obvious question: How do the critics know the premise is false? Certain methodological considerations would seem to give an advantage to the critics over those who argue in its favour. Note that we can distinguish between those statements which are verifiable (i.e. when true, can be proven true) and those which are refutable (i.e. when false, can be proven false) on purely logical grounds. Furthermore, strictly universal statements – those of the form '*all* Xs have property Y' – are refutable (if false) but not verifiable (even if true). On the other hand, strictly existential statements – those of the form 'there are *some* Xs which have property Y' – are verifiable (if true) but not refutable (even if false). At first glance it would seem that the maximization hypothesis – 'all decision-makers are maximizers' – is straightforwardly a universal statement and hence is refutable but not verifiable. But the statistical and methodological problems of empirical refutation present many difficulties. Some of them are well known but, as I shall show a little later, the logical problems are insurmountable.

The methodological problems of empirical refutations of economic theories are widely accepted. In the case of utility maximization we realize that survey reports are suspect and direct observations of the decision-making process are difficult or impossible. In this sense behavioural maximization is not directly testable. The only objective part of the maximization hypothesis is the set of logical consequences such as the uniquely determinate choices. One might thus attempt an indirect test of maximization by examining the outcomes of maximization, namely the implied pattern of observable choices based on a presumption that there is a utility function and that utility is being maximized by the choices made.

If one wishes to avoid errors in logic, an indirect test of any behavioural hypothesis which is based on a direct examination of its logical consequences must be limited to attempting refutations of one or more of the necessary conditions for the truth of the hypothesis. For example, in the case of consumer theory, whenever utility maximization is the basis of observed choices, a necessary condition is that for any given pattern of choices the 'Slutsky Theorem' must hold.[8] It might appear then that the above methodological problems of observation could be easily overcome, since the Slutsky Theorem can in principle be made to involve only observable quantities and prices. And, if one could refute the Slutsky Theorem then one could indirectly refute the maximization hypothesis.[9] Unfortunately, even if from this perspective such an indirect refutation cannot be ruled out on logical grounds alone, the methodological problems concerning observations will remain.

The fundamental methodological problem of refuting any behavioural hypothesis indirectly is that of constructing a convincing refutation. Any indirect test of the utility maximization hypothesis will be futile if it is to be based on a test of any logically derived implication (such as the Slutsky Theorem). On the one hand, everyone – even critics of maximization – will accept the theorem's logical validity. On the other hand, given the numerous constraints involved in any concrete situation, the problems of observation will be far more complex than those outlined by the standard theory. Thus, it is not difficult to see that there are numerous obstacles in the way of constructing any convincing refutation of maximization, one which would be beyond question.

I now wish to offer some different considerations about the potential refutations of the neoclassical behavioural hypothesis. I will argue here that even if one could prove that a consumer is not maximizing utility or a producer is not maximizing profit, this would not constitute a refutation of the neoclassical hypothesis. The reason why is that the actual form of the neoclassical premise is not a strictly universal statement. Properly stated, the neoclassical premise is: 'For *all* decision-makers there is *something*

they maximize.' This statement has the form which is called an incomplete 'all-and-some statement'. Incomplete all-and-some statements are neither verifiable nor refutable! As a universal statement claiming to be true for all decision-makers, it is unverifiable. But, although it is a universal statement and it should be logically possible to prove it is false when it is false (viz. by providing a counter-example) this form of universal statement cannot be so easily rejected. Any alleged counter-example is unverifiable even if true![10]

Let me be specific. Given the premise 'All consumers maximize *something*', the critic can claim to have found a consumer who is not maximizing anything. The person who assumed the premise is true can respond: 'You claim you have found a consumer who is not a maximizer but how do you *know* there is not *something* which he or she is maximizing?' In other words, the verification of the counter-example requires the refutation of a strictly existential statement; and as stated above, we all agree that one cannot refute strictly existential statements.

In summary, empirical arguments such as Simon's or Leibenstein's that deny the truth of the maximization hypothesis are no more testable than the hypothesis itself. Note well, the logical impossibility of proving or disproving the truth of any statement does not indicate anything about the truth of that statement. The neoclassical assumption of universal maximization could very well be false, but as a matter of logic we cannot expect ever to be able to prove that it is.

THE IMPORTANCE OF DISTINGUISHING BETWEEN TAUTOLOGIES AND METAPHYSICS

Some economists have charged that the maximization hypothesis should be rejected because, they argue, since the hypothesis is not testable it must then be a tautology, hence it is 'meaningless' or 'unscientific'. Although they may be correct about its testability, they are wrong about its being necessarily a tautology. Statements which are untestable are not necessarily tautologies because they may merely be metaphysical.

Distinguishing between tautologies and metaphysics

Tautologies are statements which are true by virtue of their logical form alone – that is, one cannot even conceive of how they could ever be false. For example, the statement 'I am here or I am not here' is true regardless of the meaning of the non-logical words 'I' or 'here'. There is no conceivable counter-example for this tautological statement. But the maximization hypothesis is not a tautology. It is conceivably false. Its truth or falsity is not a matter of logical form. The problem with the hypothesis is that it is treated as a metaphysical statement.

A statement which is a tautology is intrinsically a tautology. One cannot make it a non-tautology merely by being careful about how it is being used. A statement which is metaphysical is not intrinsically metaphysical. Its metaphysical status is a result of how it is used in a research programme. Metaphysical statements can be false but we may never know because they are the assumptions of a research programme which are deliberately put beyond question. Of course, a metaphysical assumption may be a tautology but that is not a necessity.

Typically, a metaphysical statement has the form of an existential statement (e.g. there is class conflict; there is a price system; there is an invisible hand; there will be a revolution; etc.). It would be an error to think that because a metaphysical existential statement is irrefutable it must also be a tautology. More important, a unanimous acceptance of the truth of any existential statement still does not mean it is a tautology.

Some theorists inadvertently create tautologies with their *ad hoc* attempts to overcome any possible informational incompleteness of their theories. For example, as an explanation, global maximization implies the adequacy of either the consumer's preferences or the consumer's theory of all conceivable bundles which in turn implies his or her acceptance of an unverifiable universal statement. Some theorists thus find global maximization uncomfortable as it expects too much of any decision-maker – but the usual reaction only makes matters worse. The maximization hypothesis is easily transformed into a tautology by limiting the premise to local maximization. Specifically, while the necessary conditions (a) and (b) are not sufficient for global maximization, they are sufficient for local maximization. If one then changes the premise to say, 'if the consumer is maximizing over the neighbourhood of the chosen bundle', one is only begging the question as to how the neighbourhood was chosen. If the neighbourhood is defined as that domain over which the rate of change of the slope of the marginal utility curve is monotonically increasing or decreasing, then at best the hypothesis is circular. But what is more important here, if one limits the premise to local maximization, one will severely limit the explanatory power or generality of the allegedly explained behaviour.[11] One would be better off maintaining one's metaphysics than creating tautologies to seal their defense.

Metaphysics vs methodology

Sixty years ago metaphysics was considered a dirty word but today most people realize that every explanation has its metaphysics. Every model or

theory is merely another attempted test of the 'robustness' of a given metaphysics. Every research programme has a foundation of given behavioural or structural assumptions. Those assumptions are implicitly ranked according to their questionability. The last assumptions on such a rank-ordered list are the metaphysics of that research programme. They can even be used to define that research programme. In the case of neoclassical economics, the maximization hypothesis plays this methodological role. Maximization is considered fundamental to everything; even an assumed equilibrium need not actually be put beyond question as disequilibrium in a market is merely a consequence of the failure of all decision-makers to maximize. Thus, those economists who put maximization beyond question cannot 'see' any disequilibria.

The research programme of neoclassical economics is the challenge of finding a neoclassical explanation for any given phenomenon – that is, whether it is possible to show that the phenomenon can be seen as a logical consequence of maximizing behaviour – thus, maximization is beyond question for the purpose of accepting the challenge.[12] The only question of substance is whether a theorist is willing to say what it would take to convince him or her that the metaphysics used failed the test. For the reasons I have given above, no logical criticism of maximization can ever convince a neoclassical theorist that there is something intrinsically wrong with the maximization hypothesis.

Whether maximization should be part of anyone's metaphysics is a methodological problem. Since maximization is part of the metaphysics, neoclassical theorists too often employ *ad hoc* methodology to deflect possible criticism; thus any criticism or defense of the maximization hypothesis must deal with neoclassical methodology rather than the truth of the hypothesis. Specifically, when criticizing any given assumption of maximization it would seem that critics need only be careful to determine whether the truth of the assumption matters. It is true that for followers of Friedman's Instrumentalism the truth of the assumption does not matter, hence for strictly methodological reasons it is futile to criticize maximization. And the reasons are quite simple. Practical success does not require true knowledge and Instrumentalism presumes that the sole objective of research in economic theory is immediate solutions to practical problems. The truth of assumptions supposedly matters to those economists who reject Friedman's Instrumentalism, but for those economists interested in developing economic theory for its own sake I have argued here that it is still futile to criticize the maximization hypothesis. There is nothing *intrinsically* wrong with the maximization hypothesis. The only problem, if there is a problem, resides in the methodological attitude of most neoclassical economists.

In summary, the general lesson to be learned here is that while it may seem useful to criticize what appear to be necessary elements of neoclassical economics, it may not be fruitful when the proponents of neoclassical economics are unwilling to accept such a line of criticism. External criticisms may be interesting for critical bystanders, but for someone interested only in attempting to see whether it is possible to develop a neoclassical model to explain some particular economic phenomenon, the questions of interest will usually only be the ones concerning particular techniques of model-building. They will usually be satisfied with minimalist concern for whether the model as a whole is testable and thus be satisfied to say that if you think you can do better with a non-neoclassical model (in particular, one which does not assume maximization), then you are quite welcome to try. When you are finished, the neoclassical economists will be willing to compare the results. Which model fits the data better? But until a viable competitor is created, the neoclassical economists will be uninterested in *a priori* discussions of the realism of assumptions which cannot be independently tested as is the case with the maximization assumption.

NOTES

1 Thus one might use Simon's argument to deny the necessity of the maximization assumption. But this denial is an indirect argument. It is also somewhat unreliable. It puts the onus on the critic to offer an equally sufficient argument that does not use maximization either explicitly or implicitly. Sometimes what might appear as a different argument can on later examination turn out to be equivalent to what it purports to replace. This is almost always the case when only one assumption is changed.

2 Note that any hypothesized utility function may already have the effects of constraints built in as is the case with the Lagrange multiplier technique.

3 This is not the error I discussed in the previous chapter, that is, the one where some people call (b) the sufficient condition.

4 Although Shackle's argument applies to the assumption of either local or global maximization, it is most telling in the case of global maximization.

5 Requiring an inductive proof of any claim to knowledge is called Inductivism. Inductivism is the view that all knowledge is logically derived generalizations that are based ultimately only on observations. The generalizations are not instantaneous but usually involve secondary assumptions which require more observations to verify these assumptions to ensure that the foundation of knowledge will be observations alone. This theory of knowledge presumes that any true claim for knowledge can be proven with singular statements of observation. Inductivism is the belief that one could actually prove that 'all swans are white' by means of observing white swans and without making any assumptions to help in the proof. It is a false theory of knowledge simply because there is no logic that can ever prove a strictly universal generality based solely on singular observations – *even when the generality is true* [see

further my 1982 book, Chapter 1].

6 Again this raises the question of the intended meaning of the maximization premise. If global maximization is the intended meaning, then the consumer must have a (theory of his or her) preference ordering over all conceivable alternative bundles. At a very minimum, the consumer must be able to distinguish between local maxima all of which satisfy both necessary conditions, (a) and (b).

7 Some people have interpreted Simon's view to be saying that the reason why decision-makers merely satisfice is that it would be 'too costly' to collect all the necessary information to determine the unique maximum. But this interpretation is inconsistent if it is a justification of assuming only 'satisficing' as it would imply *cost minimization* which of course is just the dual of utility maximization!

8 The Slutsky Theorem is about the income and substitution effects and involves an equation derived from a utility maximization model which shows that the slope of a demand curve can be *analyzed* into two basic terms. One represents the contribution of the substitution effect to the slope and the other the income effect's contribution. The equation is interpreted in such a manner that all the terms are in principle observable.

9 For example, if one could show that when the income effect is positive but the demand curve is positively sloped, then the Slutsky Theorem would be false or there is no utility maximization [see Lloyd 1965]. I will return to Lloyd's views of the testability of the Slutsky equation in Chapter 14.

10 The important point to stress here is that it is the incompleteness of the statement that causes problems. Whether one can make such statements verifiable or refutable depends on how one completes the statement. For example, if one completes the statement by appending assertions about the nature of the function being maximized (such as it being differentiable, transitive, reflexive, etc.) one can form a more complete statement that may be refutable [see Mongin 1986].

11 See note 6 above. If one interprets maximization to mean only local maximization, then the question is begged as to how a consumer has chosen between competing local maxima.

12 For these reasons the maximization hypothesis might be called the 'paradigm' according to Thomas Kuhn's view of science. But note that the existence of a paradigm or of a metaphysical statement in any research programme is not a psychological quirk of the researcher. Metaphysical statements are necessary because we cannot simultaneously explain everything. There must be some exogenous variables or some assumptions (e.g. universal statements) in every explanation whether it is scientific or not.

2 Marshall's 'Principles' and the 'element of Time'

> The Hatter was the first to break the silence. 'What day of the month is it?' he said, turning to Alice: he had taken his watch out of his pocket, and was looking at it uneasily, shaking it every now and then, holding it to his ear...
>
> 'Two days wrong!' sighed the Hatter. 'I told you butter wouldn't suit the works!' he added, looking angrily at the March Hare.
>
> 'It was the *best* butter,' the March Hare replied.
>
> Lewis Carroll

While it might not be possible to confront neoclassical theory by criticizing the maximization hypothesis, its main essential element, internal criticisms are still not ruled out. But internal criticisms of maximization are very difficult since too often utility as the objective of maximization is not directly observable. Are there any ancillary aspects of maximization that can be critically examined? Perhaps if there are, we can find them in the views that Marshall developed in his famous book *Principles of Economics* [1920/49]. Marshall, I *now* think, had a clear understanding of the limitations of what we know as neoclassical economics. Recognized limitations would seem to be a good starting point for a critical examination of neoclassical economics.

I say that I *now* have this view because as a product of the 1950s and 1960s I never learned to read originals – we were taught to be in a big hurry. Consequently I accepted the many second-hand reports which alleged that the contributions of Samuelson, Hicks, Robinson, Sraffa, Keynes, Chamberlin, Triffin and others represented major or revolutionary advances in economic science which displaced the contributions of Marshall. If the truth were told, economic theory is no better off – maybe it is even worse off.

With respect to Marshall's *Principles* the only apparent accomplishment of more modern writings is a monumental obfuscation of the problem that

Marshall's method of analysis was created to solve. A clear understanding of the methodological problem that concerned Marshall is absolutely essential for a clear understanding of the Marshallian version of neoclassical economics. Unfortunately, owing to our technically oriented training, we have lost the ability to appreciate Marshall's approach to the central problem of economic analysis which is based on the methodological role of the element of time. Having said this I do not want to lead anyone to think that I am simply saying that one can understand Marshall by mulling over each passage of everything he wrote. Reading the history of economic thought has its limitations, too. My main interest is improving my understanding of modern neoclassical economics, so I view historical works as a guide rather than a rule.[1] It is *my* understanding that is at issue, not Marshall's. Nevertheless, appreciating why Marshall saw problems with 'the element of Time' and its role in economic analysis can be a fruitful basis for a critical understanding of Marshall's version of neoclassical economics.

Unlike neo-Walrasian equilibrium models, which take time for granted, Marshall's economics allows time to play a central role.[2] Simply stated, the recognition of the element of time is Marshall's solution to the problem of explanation which all economists face. That problem can only be appreciated in relation to a specific explanatory principle or behavioural hypothesis. Such a relationship was introduced in the preface to Marshall's first edition where he refers to the Principle of Continuity. But he explains neither the role of continuity in the problem of explanation nor the problem itself. The problem, it turns out, results primarily from a second explanatory principle, the Principle of Substitution, which he introduces later (in Book V). I will argue here that Marshall saw an essential role for time in economic explanations for the simple reason that he wished to apply only these two principles to all economic problems.

THE TWO EXPLANATORY 'PRINCIPLES'

It seems surprising that there are only two explanatory principles stated by Marshall – the Principle of Substitution and the Principle of Continuity. These two explanatory principles are distinguished from 'laws' (or 'tendencies') which also play a role in his explanations. The principles are assumptions (we assume because we do not know) but Marshall considers 'laws' to be beyond doubt.

The Principle of Substitution is easily the more familiar of the two since it is merely what we now call the neoclassical maximization hypothesis. It says, *everyone is an optimizer* (i.e. a maximizer or minimizer) *given his or her situation* (including his or her endowment). But *by itself* it is not a sufficient explanation of phenomena. The Principle of Substitution presumes the truth of what Marshall calls the Principle of Continuity. Since Marshall wishes to apply the Principle of Substitution to everything, he needs to show that the Principle of Continuity applies to everything. In simple terms, the Principle of Continuity says everything is relatively a matter of degree. For Marshall there are no class differences, only matters of degree. He takes the same attitude towards the differences between 'city men' and 'ordinary people', between altruistic motives and selfish motives, between short runs and long runs, between cause and effect, between Rent and Interest, between man and his appliances, between productive and non-productive labour, between capital and non-capital, and even between needs and non-essentials. In all cases whether the degree in question is more or less is relative to how the distinction is being used *in an explanation*. For example, 'what is a short period for one problem, is a long period for another' [p. vii].[3]

Sometimes it seems that Marshall is probably the only neoclassical economist who fully appreciates the methodological problem of the applicability of the Principle of Substitution. To be sure of its applicability, he postpones its introduction until Book V, the fifth of six major parts of his book. The first four Books are devoted to convincing the reader that the assumption of maximization is applicable by demonstrating the universal applicability of the Principle of Continuity. There must be available a continuous range of options[4] over which there is free choice (i.e. substitutability is precluded whenever choice is completely limited), and the choice must not be an extreme (or special) case – otherwise the question would be begged as to what determines the constraining extreme limit.

THE 'ELEMENT OF TIME'

Marshall stresses (e.g. in his original preface) that the applicability of the Principle of Continuity (and consequently the applicability of the Principle of Substitution) depends heavily on 'the element of Time'. By ignoring the element of time, our teachers (and their textbooks) would have us believe that the Principle of Substitution is the only hypothetical aspect of the 'Principles'. If one could reduce everything to maximization then explanation would certainly be made at least formally easier. Samuelson saw that it was possible for even the notion of a stable equilibrium to be reduced to the Principle of Substitution [e.g. Samuelson 1947/65, p. 5], that is, to a matter of constrained maximization. Time, if considered at all, is deemed relevant only for the proofs of the stability of equilibria. Most of us have been trained not to see any difficulty with the element of time – for

fear of being accused of incompetence.

Marshall's view is quite the contrary: the element of time is central. For instance, to presume that at any point in time a firm has chosen the best labour and capital mix presumes that time has elapsed since the relevant givens were established (viz. the technology, the prices, the market conditions, etc.), and that period of time was sufficient for the firm to vary those things over which it has control (viz. the labour hired and the capital purchased) prior to the decision or substitution. Even when its product's price has gone up the firm cannot respond immediately. Nor can it stop production and its employment of labour merely because the price has fallen [cf. p. 298]. Contrary to modern textbooks, in Marshall's economics very short-run market pressures are more 'the noise' than they are 'the signal' *when viewed from the perspective of the entrepreneur's decision process.*[5]

Time is an essential element in Marshall's method of explanation. Marshall tells us quite a lot about explanation in economics. He stresses the need to recognize the role of fixed 'conditions', but he also stresses that the 'fixity' is not independent of the defining 'time periods'.[6] Marshall's use of the term 'conditions' can lead to confusion, so it might be useful to examine his theory of explanation more specifically by distinguishing between dependent, independent and exogenous *variables*, and between fixed and exogenous *conditions*. These distinctions crucially involve the element of time.

The relationship between dependent and independent variables is supposed to be analogous to the relationship between causes and effects. Marshall, however, cautions us that all such distinctions are relative. For instance, in the very short period the market price is the dependent variable and, given the demand, the quantity supplied is the independent variable. But, in the usual short run, the market price is the independent variable and, given technology (i.e. the production function), the quantity supplied is the dependent variable.

In the preface to the *Principles* Marshall recognizes the usual type of interdependence as being an instance of the Principle of Continuity. He specifically credits Cournot with teaching us to face the difficulty of 'mutual determination'. Marshall calls this type of interdependence a mathematical conception of continuity although he refers to this conception only in regard to the relationship between causes and effects.[7] Today we might say that, in Marshall's short period, price and quantity are both endogenous variables and are *simultaneously determined* by the exogenously given technology and demand. Thus, the distinction between independent and dependent variables is only a matter of verbal convenience since both are endogenous.

Marshall regards 'conditions' as variables which are exogenously fixed during the period of time under consideration. He relies on their fixity in his explanation of behaviour where these fixed variables are the constraints in a maximization process. In this regard, Marshall's neoclassical programme is indistinguishable from the mathematical approach of his contemporary Leon Walras. However, in Walras' approach, as it is taught today, the constraints are given as stocks to be allocated between competing uses. And, of course, Walras is usually thought to consider all processes to be completed simultaneously as if the economy were a system of simultaneous equations. Nevertheless, although both approaches to explanation are 'scientific' in Marshall's sense, the mathematical conception of an economy is rejected [p. 297].

In Marshall's view the problem of explanation is that there are too many conceivable 'causes'. It is not that one has to rely on exogenous givens as being 'causes' in any hypothesized relationship, but rather that there are so many exogenous variables to consider. This problem was not the one faced by followers of Walras who are more concerned with the solvability of his system of equations. Marshall's problem was the direct result of the method he used to deal with the necessity of conditional explanations. Where followers of Walras in effect try to attain the greatest generality or scope of the explanations by maximizing the number of endogenous variables and minimizing the number of exogenous variables, Marshall deliberately adopts a different strategy by attempting to maximize the number of fixed exogenous variables at the beginning of his analysis so as to reduce the explanation to a sequence of single-variable maximizing choices. All other variables are fixed because they are exogenous givens or because they are exogenously fixed by a prior maximization process. The exogenous reason that they are fixed in any problem is the logical basis for their use in his explanation.

There is a difficulty with Marshall's approach to explanation whenever there are many variables. It is difficult to distinguish between the endogenous conditions – those which are exogenously fixed for the period of time considered (e.g. fixed capital in the 'short run') – and the truly exogenous conditions that can never be explained as outcomes of a maximization process (e.g. weather, social conditions, states of knowledge, etc.). Although exogenous variables need not be fixed, in Marshall's approach they are treated as fixed by limiting the length of the period of time to which the explanation refers.

In Marshall's view, the problem of explanation is thus one of carefully defining the fixity of the 'conditions' by defining the relevant period of time for the operation of the explanatory Principle of Substitution. Of course, what is a relevant period of time depends conversely on what are

the relevant exogenous conditions for the application of the Principle of Substitution. For example, in Marshall's short period – 'a few months' [p. 314] – virtually everything but the level of output and the amount of labour employed is by definition fixed; but in his long period – 'several years' [p. 315] – everything but technology and social conditions is endogenous.

As with Walras' economics, in Marshall's economics the truly exogenous variables are the only bases for explanations. Any variable which is fixed for a period of time and which serves as a constraint on anyone's maximization process must be explained at some stage or be explicitly identified as an exogenous variable. More important, if it is not an exogenous variable, its fixity at any stage must be explained in terms of acceptable exogenous variables.[8] Even though Marshall's approach begins by maximizing the number of fixed exogenous variables, his ultimate objective is, like that of the followers of Walras, to explain as much as possible. Since by definition exogenous variables are those which are to be left unexplained, the Marshallian methodological strategy then is to reduce the number of exogenous variables in stages. Marshall obviously considered the methodological problem of explanation in economics to be solvable.

In Marshall's economics the truly exogenous variables are the only 'causes' in the strict sense. According to Marshall's view, if one is to provide a long-run explanation, 'time must be allowed for causes to produce their effects' [p. 30]. Of course, this 'is a source of great difficulty in economics [because] the causes themselves may have changed' [p. 30]. Note, however, that the changeability of 'causes', that is, the changeability of exogenous variables, is *not* the problem of explanation, but rather, it is the more narrow methodological problem of *verifying or refuting* one's explanation.[9]

Even when changes in the exogenous givens are assumed away, the fundamental problem for all explanations involving time still exists. The logic of explanation (for example, of all the co-determined endogenous variables) requires that we recognize at least one exogenous variable; and given maximization with exogenous tastes and exogenous constraints, changes in endogenous variables are explained as being caused by changes in at least one of the exogenous variables. But this means that an explanation of long-run dynamic behaviour requires at least one exogenous variable which is impervious to the amount of real time elapsed in the long run (otherwise, the explanation might be circular). For this purpose, the explanatory element of time involves the identification of at least one time-independent exogenous variable – that is, one which does not change over the defined long run.

It should be noted that Marshall's view of explanation also recognizes another aspect of the element of time. *If* the state of affairs at any point in time is to be *explained* as a consequence of someone's optimizing choice, it *must have been* possible to alter one's choices – and this possibility is both a matter of the time available and the continuity of options. Needless to say, it also presumes the ability to know what is the best *option*. Learning what is the best option takes time [p. 284]. This question of learning, I would argue, is *the* explanatory problem involving the element of time. Of course, for Marshall, the inductive scientist, time is all that is necessary for the accumulation of the needed knowledge. Unlike the classical school, Marshall sees no need to assume 'perfect knowledge' because he explicitly wishes to recognize the period of time under consideration – a period he would consider sufficiently long to obtain any 'necessary knowledge'.[10]

MARSHALL'S STRATEGY

It would be misleading to suggest that Marshall's problem of explanation is merely a matter of defining a long-run equilibrium, for it is also a matter of how the long-run equilibrium is reached. Again, in Marshall's view [p. 304], the explanatory problem is that there are too many exogenous variables in the short run during which most decisions are made. His strategy is intended to reduce the number of exogenous variables by increasing the number of variables to which the Principle of Substitution can be applied at later stages.[11] Marshall thus considers the problem of explanation to be solvable since he recognizes that there is a different degree of changeability for each variable (another application of the Principle of Continuity). In short, Marshall's strategy is to distinguish between short-run and long-run explanations. Any complete explanation must specifically *assume* which variables can be changed most quickly – that is, the variables must be ordered according to their changeability. Different orderings may yield a different path to the long-run equilibrium. Unless the assumption is very specific it may be impossible to distinguish between a long-run moving equilibrium and a short-run movement toward a new long-run equilibrium.

Although Marshall gives a prominent role to the distinction between long and short periods, it is not sufficient to solve his problem of explanation – which, as I have said, is a problem concerning the methodological choice of exogenous variables that are impervious to time. Yet most commentators seem to think that Marshall's 'statical method' – namely, the contents of Book V – constitutes his solution to the problem of explanation. This is a mistake.

The first point to be made is that Marshall's 'statical' or partial equilibrium method of analysis yields incomplete explanations. The

'statical' method is relevant only for decisions 'on the margin' or in the neighbourhood of an equilibrium position. By itself the method examines the necessary but not the sufficient conditions for equilibrium. The second point to be made is that Marshall does offer a more complete explanation which is based on the contents of Book IV. By itself, Book V deals only with the 'noise' in order at best to explain it away. A source of an explanation of an economy's true dynamics and its application of the Principle of Continuity to the element of time is to be found in Book IV. These two points will be discussed in turn.

The insufficiency of Book V

I do not think Marshall ever claims that Book V alone represents a complete explanation of an economy's behaviour. Yet, judging by modern textbooks, one could easily think that Book V is 'the principles of economics'. What we call microeconomic analysis today can all be found in Book V. Nevertheless, implicitly Book V provides only the necessary conditions for any equilibrium. That is, on the *assumption* that an economy is in long-run equilibrium at a point in time, certain necessary relationships must hold whenever that assumption is true. It is a 'statical' method because it may be relevant only for that one equilibrium position at one point in time. In effect, Book V examines the local stability properties of the *assumed* long-run equilibrium that are the logical consequences of definitions of equilibrium and of the long period. But it will be argued below that the stability properties are heavily dependent on the empirical assertions of Book IV.

To be specific, before Book V can be considered relevant for anything, that is, before it can play a role in economic analysis, a key question must be asked: why should there ever be a long-run equilibrium? Marshall approaches this question in two ways. The most familiar is in Book V where he defines an ordering of the changeability of the variables with respect to three periods of time – 'the very short period', 'the short period' and 'the long period'. The quickest variable in Marshall's world is the market-determined price. In fact, his definition of a market is not the textbook one of a *place* where buyers and sellers meet to haggle over the price. Marshall makes the existence of a market depend on whether the price clears *quickly* enough for all producers to face the *same price* regardless of their location. For Marshall then there is no market for any good whose price is either not uniform[12] or not quickly established. In effect, this axiom about market prices makes all firms price-takers since it takes longer to establish their (short-run) decisions than the price itself.

Marshall's definition of the market means that the market price (as opposed to the short-run or long-run equilibrium price) is the only *real time* observable price. This theory of market prices assumes that the supply quantity is fixed – virtually everything is fixed but the price. The remainder of the discussion in Book V is an examination of what happens to the market price over time when more and more of the fixed givens are allowed to change. For example, Marshall begins by allowing the firms to make substitutions in their quantity supplied in response to the current level of the market price (relative to costs). This 'short-run' *process* of substitution requires some time – 'a few months or a year' [p. 314].

Marshall says that he wishes to argue that demand determines the market price in one extreme – the very short run – and technology determines the market price in the other extreme – the long-run equilibrium. Implicitly the real world is somewhere in between.[13] Again, the meaning of 'determines' is only a matter of relationships made *necessary* by virtue of his defined equilibria. If at a point in time the economy is at a long-run equilibrium, it must also be at a short-run equilibrium, since if it were not there would be short-run incentives to change the givens which are the constraints in the determination of the market price. Similarly, the short-run equilibrium presumes that the market is in equilibrium. In other words, every long-run equilibrium must also be a short-run equilibrium and every short-run equilibrium must be a market-run equilibrium. This 'nesting' of the forms of equilibrium is the essence of Marshall's 'statical method'.

Although it is now very easy to list the *necessary* conditions for the existence of a long-run equilibrium, the key question still concerns the *sufficient* conditions for the existence of a long-run equilibrium, which must be consistent with both a short-run equilibrium and a market equilibrium. The question of consistency has been a major source of controversy over the last sixty years. The logical problem is that the absence of excess profits in conjunction with profit maximization in the long period implies that the production function is locally linear-homogeneous (constant returns to scale on the margin); but this implication appears to be inconsistent with a downward sloping demand curve, the ultimate constraint thought to be necessary to limit the size of the producer.[14]

Marshall's only line of defense is his other approach, which is based on the Principle of Continuity. Given the continuous operation of the Principle of Substitution, it is quite possible for the price to be above or below the long-run equilibrium price. When it is above there are positive excess profits and when it is below there are losses and, logically, there must be a (long-run equilibrium) point in between where excess profits are zero. The apparent inconsistency is due only to the discussion of the hypothetical and

heuristic 'stationary state' – it is a very special type of long-run equilibrium which is supposed to hold for a specified period of time. The only inconsistency is between the previously mentioned nesting of equilibria and the stationary state. Specifically, the inconsistency is that the stability of each of the various equilibria that hold at the long-run equilibrium depends necessarily on the consideration of *different* periods or lengths of time for each whereas in the stationary state they are all supposed to refer to the *same* period of time.

Leaving the stationary state aside, there is no reason why the stability of the various forms of equilibrium has to refer to the same set of 'conditions' or variables or, equivalently, to the same period of time. Hence, the stability relations (e.g. the necessary slopes of curves) for one form of equilibrium will not be 'statically' consistent with those relations necessary for the stability of another form. If one ignores the element of time, it is only too easy to 'see' an inconsistency where otherwise there is none.

The methodology of Book V vs a complete explanation

Once one recognizes the necessary element of time it might appear that there is no logical problem with Book V. But to the contrary, there still remains the matter of explaining *why* there should ever be a long-run equilibrium,[15] and this is a question which must be tackled within an appropriate frame of reference. The essential element of the frame of reference of any behavioural explanation is the specification of exogenous and endogenous variables. All explanations must be based on something being exogenous. In Marshall's time-based view of the economy, it must be something whose exogeneity extends to a longer period of time than the 'long period' under consideration. Marshall deals with this issue first in Book IV.

Particularly relevant to Marshall's explanation of an economy is what is sometimes called his 'life-cycle' hypothesis of the firm. In its most specific form it is an empirical assertion about the history of an individual firm with a life-span of three generations [cf. Hague 1958; Loasby 1978]. In its more general form it says that at the beginning of its life the firm benefits from learning so that its ability to produce increases with its size. Implicitly Marshall is only concerned with growing firms – their size is irreversible, hence time and size go together. At the end of its life every firm suffers from diminishing returns. In either case, the life-cycle trajectory is the needed long-run exogenous variable which provides the essential frame of reference.

By itself, this hypothesis about the beginning and the end of the life of a firm does not seem very relevant. The addition of the Principle of Continuity, however, renders the desired result. This principle allows us to conclude that, since returns change from increasing to decreasing, at some point in between there must have been 'constant' returns. This point is a *possible* long-run equilibrium. Given the life-cycle hypothesis and continuity, every firm must pass through this point. Once it is reached, the 'statical method' can be used; but it remains merely a 'snapshot', relevant only for that one point (in the history of the firm).

There is absolutely no reason why all the firms in an economy should simultaneously reach the point of constant returns – that is, reach the 'turning point,' as Marshall calls it. It might be interesting for someone to explore such a fantasy world, but nowhere does Marshall seem to be suggesting that such a state of affairs is *necessary*. Book V nevertheless explores the nature of this turning point: Book V 'is not descriptive, nor does it deal constructively with real problems' [p. 269]. However, Marshall does say Book V 'sets out the *theoretical backbone* of our knowledge of the causes which govern value' [pp. 269–70, emphasis added]. However, this statement is qualified. He says, 'it aims not so much at the attainment of knowledge [but rather] at the power to obtain and *arrange* knowledge with regard to two opposing sets of forces' [p. 270, emphasis added].

Marshall's use of the words 'theoretical' and 'arrange' differs slightly from the usual modern usage. His usage is related to Milton Friedman's *as if* approach to explanation. There is no claim that the method of analysis – of arranging the facts of business – is a true explanation. There is only the claim that the nature of the inevitable turning point can be understood to be the result *if* the world were in a state of equilibrium at a moment in time – or more properly, in a state where forces are balanced.

As in most economists' adventures in methodology, Marshall wishes to be all things to all people; thus his is not a pure example of the Instrumentalism we associate with Friedman.[16] Rather, the Introduction to Book V gives a classic example of what we now call Conventionalist methodology. We are offered *a way of looking at things*. What is offered is not claimed to be true; it can be judged only to the extent that it is *better* or *worse* than some other competing view. Book V is filled with conventions with no claim to their truth status (e.g. the representative firm, the stationary state, the market, the long period, etc.). Only in those cases where we know that he thinks a particular convention is a fiction do we have examples of the 'as if' methodology.

The methodology discussions of the *Principles* are not very interesting today but his theory of the firm should be. The point at issue is that Book IV is a foundation for a *complete* theory of the firm: the firm is always to be found somewhere on its life-cycle trajectory. Its location on the trajectory is determined completely by the time elapsed, [cf. p. 258], but

the value of that position can only be determined as a relative value, relative to its past and its future. There are simply too many contingencies to be able to determine the absolute value. But remember, the Principle of Continuity is only concerned with relative values.

Book V does offer a way of seeing the absolute value as a consequence of external forces, that is, of competitive market pressures. But there is no reason why the actual, real-time values would ever be 'long-period normal' prices. The existence of long-period normal prices is merely, one might suggest, a beautiful fiction which lends itself to simple mathematical analysis having no bearing on 'real problems' [cf. p. 269].

INADEQUACIES OF MARSHALL'S METHOD VS PROBLEMS CREATED BY HIS FOLLOWERS

Over the last sixty years there have been two major problems in the application of Marshall's principles; both of them involve the element of time. The first concerns the meaning of increasing returns and the nature of the long-run equilibrium. The second concerns the artificial distinction between 'historical' and 'logical' time.

Problems with the firm's long-run equilibrium

Marshall's Victorian style lends itself easily to distortion. What he meant by certain words in one place may not have the same meaning in another. For example, the term 'increasing returns' is used in two different senses; both result from his implicit assumption that the firm is always growing; hence size and time go together. In Book V he uses the term to describe the observation that average productivity rises over time for any given input levels [p. 377]. This use is at variance with modern usage. Earlier, in Book IV, he employs the term in the limited modern sense to mean an increase in output which is proportionally greater than the increase in the size of the firm [p. 266]. A similar confusion derives from his use of the term 'margin' when discussing his 'representative firm'. By definition, the representative firm is at the 'turning point' on the life-cycle trajectory. At that point average and marginal cost both equal price; thus it is possible to use the average and marginal magnitudes interchangeably. But another use of the term 'marginal' emerges when he refers to the representative firm's contribution to its industry's output.

These confusions are merely irritants. The major problem is the one which occurs when critics ignore the element of time inherent in the 'statical method' whenever that method is applied to long-run equilibria (as noted above). Although the difficulty is primarily logical, it results from conjoining four statements whose individual truth status depends on different periods of time. They are the following:

(a) Prices are determined before the firm makes its supply choice; hence prices are given.

(b) The *Principle of Continuity* applied to all inputs (all inputs are variable) means that the production function of the firm is locally linear-homogeneous and that the level of output is always equal to the sum of the marginal productivities, each multiplied by the respective input (Euler's theorem).

(c) The *Principle of Substitution* (i.e. profit maximization) applied to all variable inputs means that the marginal productivity of each input multiplied by the product's price will always equal the price of that input.

(d) The firm is at the 'turning point', that is, its excess profits are zero.

There is no difficulty with the conjunction of these four statements if they only refer to a single point in time.[17] Moreover, even over the short run, given statement (a) any two of the remaining statements imply the other one.[18] So long as the theory of the firm is confined to the 'short period' there need not be any logical problems. The problems that are alleged to exist arise only when the theory (i.e. the Principle of Substitution) is applied in the long-run period to the short-run *constraints*.

Applications of the Principle of Substitution involve some form of maximization (or minimization) facing fixed constraints. In the short run, all the variables which (by definition) cannot be varied constitute the short-run constraints (e.g. the short run may presume capital is fixed while labour is variable). In the long run everything except the production function is supposed to be variable (by definition); but this raises a major methodological problem. Anything which is variable must logically be subjected to the Principle of Substitution. This means that the variables that served as fixed constraints in the short run become endogenous variables in the long run. But this also means that there are no constraints in the long run and this leaves the Principle of Substitution inoperable in the long run. In the long period, then, the conjunction of the assumptions of a price-taker, (a), of the changeability of all variables in the production function, (b), and of profit maximization with regard to all changeable variables, (c), seems to deny any limit to the size of the individual firm – as if size has nothing to do with time (this interpretation of Marshall's theory of the firm, by its focusing only on the internal logic of maximization, is quite contrary to the views expressed in Book IV).

The methodological problem of explaining the size of the firm (as a

consequence of maximization) seems to have troubled many of Marshall's followers although it did not seem to trouble him since his Principle of Continuity discourages extreme viewpoints, such as long-run equilibria. The problem only arises when one attempts to apply the Principle of Substitution to the size of the firm *in the long run*. Today this problem is avoided (i.e. swept under the rug) by saying that one should only *explain* the size of the industry. But this tactic merely raises other questions such as What prevents any one firm from taking over the industry as a monopoly?

Although there is considerable discussion of industries in the *Principles*, Marshall's explanatory Principle of Substitution is applied only to the (short-run) decisions of the individual firm. The industry is merely an epiphenomenon – the logical consequence of what all individual firms do. This is a standard neoclassical viewpoint. However, this viewpoint has always posed certain puzzles concerning the interaction of demand and supply in the market. The difficulty is that both the market and the industry are defined for a specific good but the market is related to the individual firm only through the going price. The price by itself says nothing about quantities except that aggregate quantity demanded must equal industry supply. But, if individual firms must determine the quantity supplied independently of each other, the aggregate quantity supplied is only an epiphenomenon. In terms of Marshall's individualistic methodology, this approach to the relationship between firm and industry appears rather mysterious.

To overcome the mystery, Marshall offers the infamous heuristic fiction, the representative firm. Unfortunately, whenever one tries to use the representative firm, instead of Book IV, to explain the size of the firm as just another consequence of an application of the Principle of Substitution, another methodological problem is created. Recall that the representative firm is defined [p. 285] as a firm at the 'turning point' and it is also a firm on the margin of the industry (older firms will be making less than normal profits). As a profit maximizer at the turning point (where profits are just normal), the representative firm must face constant returns to scale (at least 'locally' [see Baumol 1977, p. 578]). On the other hand, as a representative of the industry, it must be constrained by the negatively sloped demand curve. This latter constraint means that we have a fifth statement which must be conjoined with the other four, namely:

(e) The representative firm's marginal revenue must be less than the price.

The problem is that either statements (e) and (a) are mutually contradictory or one of the other statements must be denied. With respect to any one firm it is not possible for all five statements to be true simultaneously. For example, while profit maximization implies the equality of marginal cost and marginal revenue, zero excess profits implies an equality between average cost and average revenue (the price). Thus, when marginal revenue is less than the price, the firm must be operating where there are increasing returns (since marginal cost must be less than average cost) which is contrary to statement (b). Note that a firm can still be a price-taker even when its average revenue is falling with the quantity supplied.

It could be speculated that all of the controversies surrounding the long-run theory of the individual firm are merely about which of the five statements should be dropped.[19] Moreover, most of the controversies have ignored the element of time. There is no doubt that *if* one ignores the element of time (which differs according to the statement one is considering) and, instead, views the above statements as holding at a single (static) point of time, then logically some of the statements are mutually inconsistent. As argued by Piero Sraffa [1926] and Joan Robinson [1933/69], something must give. A realistic interpretation is that the idea of a price-taker, (a), must go, but Marshall's statical method of dealing with his problem of explanation – distinguishing between very short periods and the short run – blocks that avenue. Allowing that prices may not be market-determined would lead to a conclusion that is contrary to Marshall's objective. If prices were *not* determined in a market, then demand could only play a role in the determination of the size of the *industry* – that is, given the life-cycle, demand determines the number of firms in an industry – *in the long run*. Prices are left to be determined by technical and social considerations within and between firms (e.g. without 'spoiling the market' [p. 313]).

Today, such conclusions seem to be ideologically unacceptable or mathematically inconvenient for economic theorists – hence we simply have stopped talking about Marshallian economics since what he promised (namely, a role for demand and utility maximization in the determination of prices) seems doomed. What I am suggesting here is that things may not be as desperate as everyone seems to fear. Perhaps all that is required is a proper examination of the *element of time*.

The distinction between logical and historical time

Contrary to Marshall's view, it is claimed by post-Keynesians that one must carefully distinguish between 'historical' and 'logical' time [e.g. Robinson 1974]. Historical time refers to the usual calendar or clock time within which decision processes are irreversible. In logical time decisions are reversible. For example, the life-cycle hypothesis is in historical time since it is assumed that the firm always gets older; it cannot get younger.

One might say that this is because with the passage of time the firm is learning but it cannot 'unlearn'. The stability analysis of equilibrium theory is in logical time since the analysis is always conducted in terms of questions such as What *if* the price were higher or lower than the equilibrium price? Logical time is concerned with conceivably possible alternative worlds (regardless of actual events) at any given point in time, whereas historical time may be concerned with the (necessarily) singular event occurring at that time and the accumulation of learning which has transpired up to that point.

The distinction between historical and logical time corresponds respectively to Books IV and V. But the intellectual separation of these concepts (and Books) into mutually exclusive classes is a direct contradiction of Marshall's Principle of Continuity. Marshall does not claim that these concepts or books should be separated. To the contrary, Books IV and V go together. Reality for Marshall is on the continuum *between* the two extreme concepts, that is, reality involves both Books in full measure. Any explanation of the behaviour of an enterprise must be both grounded in history (i.e. irreversible past decisions and learning) and explanatorily complete (i.e. it must at least imply a stable determination of the values of the variables to which the Principle of Substitution has been applied).

SOME CRITICAL CONSIDERATIONS

Most of modern neoclassical economic analysis concerns only the mathematics of Book V. The reason, I think, is simply that Book V is the only part of Marshall's *Principles* that is compatible with the methodological doctrine that dominates economic theory today – Conventionalism – namely, the methodology that restricts research to questions of logical validity instead of empirical truth.[20] Economists today do not wish to discuss the 'truth' of economic theories but only examine their logical validity. The reason why logical validity rather than empirical truth is the preferred object is that with the help of mathematical analysis the former can be established more quickly. Even though Marshall stressed the importance of gradual, slow change, those economists in a hurry will find the logic or mathematics of static equilibria more interesting. Logical analysis can be very quick but real change takes real time and thus may not be disposed to conveniently easy analysis.

NOTES

1 My approach is much like Negishi's [1985]. As Negishi noted, 'What is important is not whether a particular interpretation of a past theory is correct, but whether it is useful in developing a new theory in the present' [p. 2]. Thus the onus is on me and Negishi to show that we have learned something from reading Marshall.

2 For a discussion of the problem of time in neo-Walrasian and Austrian models, see Boland [1982a, Chapter 6].

3 Unless indicated otherwise, all page references enclosed in brackets are to Marshall [1920/49] which is the eighth edition of his *Principles*, reset in 1949.

4 Specifically, there must be what modern theorists might call the 'connectedness' of choice options [see Chipman 1960].

5 The entrepreneur (or manager of the firm) must always make a judgement as to whether day-to-day changes in the market will be long-lasting enough to justify investment and hiring decisions [see p. 314].

6 Remember, according to the Principle of Continuity everything is a matter of degree.

7 His reference to Cournot has often misled modern commentators to think that the mathematical conception is all that Marshall was saying – rather than the more important methodological issue of relative degrees.

8 This is one key element in the methodological 'hidden agenda' of neoclassical economics. In neoclassical economics everything explained is seen to be the consequence of the decisions made by individuals. The explained decisions are represented by the endogenous variables in the explanatory model. The acceptable exogenous variables are limited to natural givens (i.e. to things that cannot be chosen). For more about the role of so-called methodological individualism, see Boland [1982a, Chapter 2].

9 One must be careful to distinguish between the logical validity of an explanation and the verifiability of its truth status [see Boland, 1982a, pp. 102–4 and Chapter 1].

10 See note 5 of Chapter 1. For more on the role of inductivism in economics, see Boland [1982a, Chapters 1 and 4].

11 The variables to be treated later, then, are 'independent' variables.

12 Marshall allows for price differences that result from transportation costs [p. 271].

13 That is, the very short run is not realistic [p. 304], and the logical consequence of a long-run equilibrium is a stationary state [p. 315, footnote 1]; but a stationary state is alleged to be 'a fiction' [p. 305].

14 I will discuss Marshallian models of the firm which try to accommodate downward sloping demand curves in Chapter 5. For a different discussion, see Boland [1986a, pp. 25–8].

15 Book V discusses only the logical possibility of a long-run equilibrium.

16 For a discussion of the Instrumentalism associated with Friedman, see Boland [1982a, Chapter 9].

17 For a more detailed discussion of the question of time in neoclassical economic theory, see Boland [1982a, pp. 97–8].

18 I will examine this relationship between these statements much further in Chapter 5.

19 This is a speculation to be explored more fully in Chapter 5.

20 Conventionalism is the defeatist doctrine based on the recognition that an inductive proof is impossible. The Conventionalist alternative to inductive proofs is to prove something else. Rather than look for a proof of the one true theory, Conventionalism would have us choose the best theory recognizing that the best may not be true (as I noted earlier in this chapter). See further, Agassi [1963], Tarascio and Caldwell [1979] and Boland [1982a, Chapters 7 and 8].

3 Marshall's 'Principle of Continuity'

> If the book has any special character of its own, that may perhaps be said to lie in the prominence which it gives to ... applications of the Principle of Continuity.
>
> Alfred Marshall [1920/49, p. vi]

Neoclassical economics is primarily a method of analysis. It is the method of explaining all behaviour as the logical consequences of one behavioural assumption – namely, maximization subject to explicit constraints.[1] But, many critics ask, is the maximization hypothesis a sufficient basis for neoclassical economics? We saw in the previous chapter that according to Marshall the use of the neoclassical maximization hypothesis necessarily depends on what he called the Principle of Continuity. Contrary to the modern preoccupation with Marshall's Principle of Substitution (in the form of the neoclassical maximization hypothesis), in the first preface to his *Principles* Marshall clearly indicates that he gives primacy to the other principle. If the Principle of Continuity is so important, clearly it must be a fertile ground for critical study. For this reason it is important to understand what Marshall meant by his Principle of Continuity and why he thought it was so important.

The obvious reason for giving prominence to the relatively unknown Principle of Continuity is that the continuity of the domain of the maximization function is a *necessary condition* for *application* of the usual assumption of maximizing behaviour.[2] And even though continuity is necessary, too often it is taken for granted. Thus, Marshall rightfully devotes most of his *Principles* to an examination of the nature of an economy to determine when the Principle of Continuity can be applied. And for those circumstances where it is applicable, he devises an admittedly 'unrealistic', mechanical method of overcoming the problem of its necessity. This is his 'statical method' which I discussed in Chapter 2. The objective of this chapter is a critical examination of the methodological

presumption of continuity. Since Marshall so strongly emphasizes continuity, it is important that his method of assuring its applicability be understood.[3]

MARSHALL'S PRINCIPLE OF CONTINUITY AND HIS BIOLOGICAL PERSPECTIVE

The non-mathematical version of the application of the Principle of Continuity was very popular at the end of the nineteenth century – especially among *aficionados* of biology. But Marshall wishes to go far beyond biology. He attempts to apply this principle to everything by showing that everything is a matter of degree. Modern axiomatic model-builders discuss a form of the Principle of Continuity which is considered a question of the 'connectedness' of choice options [e.g. see Chipman 1960]. Specifically, the range of possible choice options must be continuous even when the continuum is subdivided into finite sets of categories (with no gaps or empty categories). Discreteness of choice options does not imply a non-continuity. Even when one defines the choice set as a finite set of discrete (or lumpy) options, the discreteness of the options must have been defined over a continuous background range.[4] That is, what we call a discrete point will be defined in terms of one or more continuous dimensions such that the point is located at one distinct location on a continuum. In short, it is impossible to avoid continuity, thus the only question of applicability is whether there are external limits (constraints) on the choice set.

While the relatively unknown Book IV of Marshall's *Principles* is seldom discussed today, it is central since it is devoted almost exclusively to the question of whether one can truthfully assume the applicability of the Principle of Continuity. Marshall's objective is to establish one of the primary conditions of maximization – namely, the continuously diminishing margin. He rests the weight of his argument for continuity primarily on a foundation of biological analogies. Biology was an attractive source of analogies because in Marshall's day it was seen primarily as the study of slow, gradual and progressive change along a continuum. In many cases, Marshall's argument for continuity of a variable rests only on an observation that the variable can be changed in degrees. He refers to 'man's power of altering the character of the soil' [p. 122]; and he often discusses growth: Growth of Population [Chapter 4], and of Wealth [Chapter 7]. Although growth can be distinguished from development, development usually depends on growth, thus Marshall devotes most of Book IV to the consideration of the development of a growing enterprise. The continuum that Marshall wishes to establish concerns the 'division of labour'.

It was apparently well known that 'organization increases efficiency' [p. 200]. For nineteenth-century economists, the key to this 'biological doctrine', whenever it applies to economics, was the recognition that the growth of an organization goes hand-in-hand with an increasing division among its functions – which can be viewed as either increasing disaggregation or decentralization, so to speak, or as breaking down into smaller and more specialized functions. But the more specialized (and hence decentralized) a functional part becomes, the greater the need for organization to keep all the functional parts coordinated and cooperative. The growth of an industrial organization was seen in these terms. But Marshall recognizes that there were certain drawbacks to increasing organization.

While initially the increasing organization facilitates a division of labour and its resulting economies, eventually the size of the organization reaches a limit where, given the size of the market, further growth or development of the organization tends to reduce the effectiveness of the organization. Thus, Marshall can see a life-cycle continuum which goes from increasing returns to decreasing returns. This proposition – the inevitability of decreasing returns as size increases – is considered to be true by analogy with biological systems. Marshall's objective, however, is to establish both the continuity of (average) returns and the fact that the (average) returns must eventually diminish. Once that objective is reached, Marshall has, in effect, shown that since an average cannot go from increasing to decreasing without a fall in the margin, marginal returns must be diminishing with regard to the extent of organizational development.

A necessary condition for maximization of a function over the domain of a given variable is that the value of the first derivative (i.e. the margin) be falling at the point of the maximum. In Book IV, Marshall establishes the continuity and the necessity of a maximum by means of biological analogies. With such analogies he also establishes the necessity (the 'law') of diminishing marginal productivity in the supply of all goods. It should be noted that Marshall has little difficulty in establishing the corresponding law of diminishing marginal utility. Marshall simply asserts in Book III that there are continual 'gradations of consumers' demand' [Chapter 3] and that obviously all wants must be satiable – that is, for any good there is a quantity at which utility is maximum. Thus the result is obtained that if total utility can go continuously from zero to a positive value and back toward zero, average utility must eventually fall with increasing consumption. By the same mathematical argument that is used for productivity, whenever the average is falling the marginal must be less than the average. Thus, specifically, marginal utility must (eventually) be falling since eventually average utility must fall.

Marshall thus establishes to his satisfaction that every theory that has anything to do with demand or supply must involve 'continuous gradations'. Furthermore, by adding his life-cycle theory of the firm and his assertion that all wants are satiable, he has completed the foundation (i.e. the necessary conditions) for his programme of economic analysis.

MARSHALL'S PRINCIPLE OF SUBSTITUTION AS A RESEARCH PROGRAMME

It would appear then that, once the Principle of Continuity is applied and the appropriate diminishing margins are established, the way is clear for a direct application of the Principle of Substitution to all decisions concerning demand or supply. But as I noted in Chapter 2, Marshall claims to the contrary; there are difficulties with the 'element of Time' [pp. 92 and 274]. The difficulties, however, lie in his conception of the essence of 'scientific' explanation – namely, the notion of cause and effect relations. The problem with economic explanations, according to Marshall, is that at any point of time there are too many exogenous conditions to consider. Thus he claims that all 'scientific' explanations are conditional – in particular, they depend on the assumptions made about the relevant exogenous variables. Changes are explained only as the effects of changed conditions.

Again, unless the changeability (or fixity) of the 'conditions' is explained, the Marshallian method of explanation runs the risk of profound circularity. Circularity might be avoided by adopting the Walrasian approach, but doing so would only risk an infinite regress.[5] Moreover, the completion of the Walrasian programme of representing the economy with a set of simultaneous equations turns out to depend intimately on the mathematical form of those equations. Thus, where Marshall's programme runs the risk of circularity, Walras' programme runs the more obvious risk of arbitrariness if one does not attempt to explain one's choice of hypothesized mathematical forms.

MARSHALL'S REJECTION OF MECHANICS AND PSYCHOLOGY

Summarized this way, Marshall's research programme sounds rather mechanical. Marshall states that he wishes to avoid identifying economics with the immutable laws of physics [p. 37]. Yet he thinks economics can be more rigorous and less subjective than the 'scientific' study of history. In effect, he sees biology as an intermediate stage on a continuum between inexact, subjective historical studies at the one extreme and precise, objective physics at the other extreme. Thus he draws parallels between economics and biology by seeing them both as studies of growth and development of organisms or organizations. The mutability of the character and purpose of individuals and groups ('races') of individuals in response to changing conditions is the key to the parallels. He says the same must be true for economic analysis [pp. 30–1].

Many writers, such as G.F. Shove [1942], have noted Marshall's apparent love for biological analogies. But why was Marshall so enamoured of biological analogies? Marshall's advocacy of a biological perspective in economics appears to be due to the prevailing dissatisfaction with both the mechanics of physical analogies and the dreaded 'hedonism' implied by basing economics on the psychology of the individual.

Marshall's use of biological analogies can be better appreciated when it is contrasted with the prevailing public opinion at the time he began work on his *Principles*. Prior to the French Revolution at the end of the eighteenth century, most intellectuals on both sides of the Atlantic were convinced that the apparent success of Newtonian mechanics demonstrated the correct approach to solving all social problems. Namely, if everyone were 'rational' like the scientists, they would all see that the solution to the eighteenth century problem was the elimination of both the monarchy and the Church. This revolutionary social programme collapsed in Europe with the failures of the French Revolution. Although in many ways this programme lived on in the economic principles of the Classical School as well as in the Americans' Declaration of Independence, those intellectuals disappointed with the failures of classical Rationalism hastily retreated from the objective world of 'reasonable men' to the Romantic worlds of subjective psychology, poetry and introspection.

In this sense it is easy to see how many intellectuals identified the classical school of economics with the failure of classical rationalism and thus economic analysis was considered suspect in many circles. The shortcomings of the subsequent Romantic view were not so apparent during most of the nineteenth century. Yet Marshall rejected Jevons' Romantic theory of value (which was based on demand rather than supply) because in Marshall's eyes this was probably seen as a retreat from one extreme (namely, exclusive mechanics of supply) to another extreme (namely, exclusive mechanics of demand). Later, Keynes, dissatisfied with Marshall's neoclassical economics, was to go all the way. In order to reject the mechanics of classical economics, Keynes endorsed a psychological basis for all businessmen's decision-making.[6] But the methodological question here is whether the rejection of mechanics necessarily entails the espousal of subjective psychology. Clearly, Marshall opted for a more liberal compromise.

A psychological basis for decision-making would seem too much like the 'immoral hedonism' often identified with the Benthamite programme of explanation where all human behaviour is considered to be the consequence of utility maximization. The major problem with psychologistic explanations is that they presume an immutable 'human nature' – for example, permanently given tastes. John Stuart Mill's *Principles* came very close to being such a theory of human behaviour. As Marshall saw this, the difficulty was not maximization, but rather the view that human nature is immutable. If human nature were immutable there would be little reason for social or economic change. To a Victorian scientist, the immutability of the human character was unthinkable. In summary, Marshall saw additional significance in the support his biological analogies gave to his discussion of continuity. He embraced biology because evolutionary biological analogies were the obvious and most palatable alternative to mechanical or hedonistic theories of economics and society.

COMPREHENSIVE MAXIMIZATION MODELS

Keynes identified Marshall with the mechanistic Classical School. Disagreement would be difficult on the sole basis of Book V of the *Principles*. But Marshall insisted that mathematical models of dynamics (and hence mechanics) would be inappropriate [pp. 382 and 637]. Nevertheless, Marshall's protestations notwithstanding, it is easy to see that all economic behavioural assumptions can be reduced to maximization (or minimization).

To see how the idea of equilibrium can be reduced to one of *universal* maximization alone, consider the two most common assumptions regarding equilibrium: (1) the assumption of the existence of a specific market equilibrium and (2) the assumption of the existence of a general competitive equilibrium. It is easy to see that both can be shown to follow from the assumption of successful maximization alone.

First, let us consider the elementary idea of a market equilibrium, that is, of the existence of a price at which demand equals supply. There are two structural elements in any market: the demand curve and the supply curve. In neoclassical economics, the demand curve is the dominant logical consequence of utility maximization in the sense that the curve is the locus of price and quantity combinations for which at any given price the indicated quantity is the total demand which results when *every* consumer is maximizing utility while facing that price. Likewise, the supply curve indicates the consequence of profit maximization where for any given price the curve indicates the total supply which is achieved when *every* firm is facing that price and is maximizing its profit. To see what it means to assume the existence of a market equilibrium whenever we are also assuming universal maximization, we need only consider the contrary implications of the non-existence of a market equilibrium. Whenever there is excess demand, some of the demanders are unable to maximize due to an insufficiency of supply at the going price. Such a disequilibrium in the market would thus deny universal maximization. And thus, when it is assumed that everyone is a maximizer, disequilibria are logically precluded.[7]

The more general assumption of the existence of a competitive equilibrium meets a similar fate simply because the assumption of a competitive equilibrium implies the absence of excess profits; that is, it implies the absence of any reason to exit one industry and enter another. It is easy to show that whenever Marshall's Principle of Continuity is applicable (such that *all* relevant factors of production are variable), total revenue must equal total costs if it is also assumed that all the factors are paid their marginal product. First, whenever a price-taking firm is maximizing its profit with respect to every factor, it must be paying each factor its marginal product. Second, whenever all factors are variable, Euler's theorem is applicable: output equals the weighted sum of all the input factors, each weighted by its respective marginal product. Putting these two considerations together, we see that whenever all factors are variable there must be constant returns to scale and thus paying factors their marginal product in order to maximize profit will exhaust the output. In other words, whenever the Principle of Continuity applies, *universal* profit maximization precludes excess profit. Thus we can see that there is no need to add an assumption which asserts the existence of a competitive equilibrium if we are already assuming universal maximization as well as assuming that all factors are variable!

These considerations would seem to lend considerable support to those neoclassical economists who, by accepting that everything reduces to the mathematics of maximization, wish to consider other territories to conquer with their maximization hypothesis [e.g. Becker 1976; Stigler and Becker 1977]. Their research programme is rather straightforward. Every decision-maker faces constraints and possesses an objective (utility) function and thus every equilibrium in society or an economy can be seen to follow from universal maximization. The theorist's task is only to describe the constraints and the objective function which is consistent with the absence of any incentive for change – that is, for example, with zero excess profit and zero marginal profit. Thus, the appearance of imperfections in competition can easily be explained away as the misperception of some economic theorists who incorrectly calculate the transaction costs of encouraging additional competition. That is, even the constraints facing all

short-run maximizers can supposedly be explained as the consequences of all individuals' maximization efforts by realistically assessing the cost of further substitutions in the constraints.

Such a programme has been applied to unusual questions such as those concerning an optimal amount of charity, an optimal marriage contract, an optimal capital punishment or deterrent, an optimal institutional environment, the optimality of being altruistic or even of voting, and so on. Of course, one is free to do or assume anything one likes, even to attempt to explain everything as an effect of maximization. Intellectual honesty, however, seems to require that all the necessary conditions of maximization must be fulfilled. One of them is the requirement of a continuity of options. By giving prominence to the Principle of Continuity (and the related 'element of Time') Marshall, to his great credit, recognized the limitations of applying the Principle of Substitution. In the absence of universal continuity and variability, Marshall implies that the assumption of maximization is not an appropriate method of analysis for all situations.

The major methodological question for proponents of neoclassical economics is 'Can maximization be the sole basis for the neoclassical research programme?'. I have argued above that the assumption of maximization alone is not sufficient; one must also assume or establish a minimum degree of continuity. For those who wish to extend the maximization hypothesis as a method of analysis, it is a moot point to show that the variables in question are in fact variable in both directions over a continuous range. It is all too easy to just assume that the decision-maker faces a continuum even when the choice to be made involves integer values such as when one cannot choose a half of an automobile tire or half of a radio. There are two ways to avoid this possible impasse. One could change the choice question to one involving rental time or sharing such that the choice variable more easily fits the notion of an equilibrium. Unfortunately, this type of shift in perspective usually is merely an attempt to hide the original question.[8]

Given the futility of direct criticism of the assumption of maximization behaviour, as I argued in Chapter 1, critics of the neoclassical research programme would be advised to shift their attention to the methods used (implicitly or explicitly) by neoclassical economists to establish the *applicability* of the maximization hypothesis. Surely, questions such as whether to execute a murderer or whether to vote or whether to make any irreversible decision must be a dubious territory for the method of maximization analysis. Marshall explicitly limited his analysis to those territories amenable to the Principle of Continuity. Perhaps modern 'imperialists' such as the followers of Stigler and Becker ought to learn from Marshall's avowed appreciation of the necessity of the Principle of Continuity that neoclassical models are relevant *only* if the Principle of Continuity can be shown to apply.

NOTES

1 The systematic research programme based on the universal application of maximization is the explicit methodological agenda of neoclassical economics which I discussed in Chapter 1.
2 Note that this says that it is necessary for the sufficiency of any argument employing the maximization assumption.
3 The remainder of this chapter is based on an invited paper which appeared as Boland [1990]. The copyrighted parts are reprinted here with the permission of l'Institut de Sciences Mathématiques et Économiques Appliquées and Les Presses Universitaires de Grenoble.
4 I have discussed these notions of continuity and discreteness in more detail in Boland [1986a, Chapter 5].
5 For example, to the extent that Walrasian economics is about the allocation of given resources, the question can always be begged as to where they come from.
6 I will discuss this in more detail in Chapter 9.
7 It might be argued that the stability of the equilibrium is a separate assumption, but Samuelson [1947/65, p. 5] argues that even stability conditions are formally equivalent to maximization conditions.
8 For more on this methodological strategy, see Boland [1986a, pp. 75–8].

4 Axiomatic analysis of equilibrium states

> Often mathematical formulas are used to describe certain events without awareness of the assumptions on which the applicability of the formulas depends. Even less is there thought of an investigation to determine whether the requisite assumptions are fulfilled in the real world. Therefore it is not surprising that the results are often quite unsatisfactory.
>
> On the other hand, conclusions have often been drawn from mathematical formulas, which, strictly speaking, are not conclusions at all and which at best are valid only under restrictive assumptions. The latter may not have been formulated, not to mention efforts to discover to what extent these further assumptions are fulfilled in the real world.
>
> Thus, for a fruitful application of mathematics in economics it is essential, first, that all the assumptions on which the given mathematical representation of economic phenomena depends be enumerated completely and precisely; second, that only those conclusions be drawn which are valid in the strictest sense, i.e., that if they are valid only under further assumptions, these also be formulated explicitly and precisely.
>
> If these directions are strictly adhered to, then the only objection which can be raised against a theory is that it includes assumptions which are foreign to the real world and that, as a result, the theory lacks applicability.
>
> Abraham Wald [1936/51, pp. 368–9]

> Whenever economics is used or thought about, equilibrium is a central organising idea. Chancellors devise budgets to establish some desirable equilibrium and alter exchange rates to correct 'fundamental disequilibria'. Sometimes they allow rates to 'find their equilibrium level'. For theorists the pervasiveness of the equilibrium notion hardly needs documenting.
>
> Frank Hahn [1973, p. 1]

One common avenue for criticism of neoclassical economics is to analyze the assumptions required for a state of equilibrium. Unlike the neoclassical maximization hypothesis which is deliberately put beyond question in every neoclassical model, the assumption of equilibrium is usually open to question.[1] Some models are designed to explain phenomena as equilibrium phenomena (such as prices or resource allocations). Models which offer equilibrium explanations must at least provide logically possible equilibrium states. Clearly, such equilibrium models are open to question and thus can be critically examined to determine whether a state of equilibrium is consistent with the other behavioural assumptions made. There are some equilibrium models which are not easily criticized such as those which put the existence of equilibria beyond question (e.g. those which involve the Coase theorem or unobserved transaction costs). These necessary-equilibrium models are most often used to explain away alleged disequilibrium phenomena (e.g. involuntary unemployment or socially unacceptable levels of pollution).

In this chapter I will be concerned only with models that explicitly claim to offer explanations in which it is asserted that the phenomena in question are equilibrium phenomena. In the next chapter the focus will be models which by claiming that the phenomena are disequilibrium phenomena posit the equilibrium state as an unattainable ideal.

Equilibrium models which explain why the phenomena occur usually do so by stating a series of explicit assumptions which together logically entail statements representing the phenomena in question. Now, the most common models are ones which represent each assumption with an equation and thus show that the solution of the system of equations is a statement representing the phenomena. Where there is a solution there must be a problem (except perhaps in chemistry). In this case the problem is to find values for the endogenous variables which (given the values of the exogenous variables) allow all the assumptions to be *simultaneously* true. There may be many sets of such values. When there is just one, we call it a *unique* solution. If none is possible we say the model is unsolvable. If one could never solve the system of equations, then the model cannot explain the phenomena as equilibrium phenomena.

When do we know that we are successful in explaining something? There are two necessary conditions. The first is the easiest. Most economists seem to agree that we are successful when the theory we construct is shown to be internally *consistent* and is shown to allow for the *possibility* of the phenomena, that is, when the theory does not contradict the phenomena to be explained. If we look closer at the notion of explanation we will find that this consistency criterion for success is insufficient. The condition that causes difficulty is the second one. Specifically, if one is to explain why prices are *what they are* then for a *complete* explanation (i.e. beyond just possibilities) one must also explain

why prices are not *what they are not*. In this chapter I shall examine these two necessary conditions of a successful explanation. Namely, I shall examine why we are successful in explaining any particular phenomena *only* when our theory is not only consistent but is also 'complete' with respect to those phenomena.

ANALYZING THE LOGICAL STRUCTURES IN ECONOMICS

Analyzing the success or failure of logical structures such as equilibrium models is not a new enterprise. Indeed, for a long time it has been an interest of pure mathematicians and some mathematical economists who engage in what they call axiomatic analysis or axiomatics.[2] Their efforts have been directed only at the formalistic aspects of logical structures and thus they have too often been more concerned with axioms of *language* models where the *form* of the axiomatic structure remains the same and the interpretations of the axioms differ to produce various languages [e.g. see Koopmans 1957]. I think axiomatics can also be of considerable importance for our critical understanding of economic phenomena. The primary importance of axiomatics is that it can offer a means of systematically criticizing a given theory (i.e. a given set of assumptions).

For the purpose of critical understanding, the two primary tools of axiomatics are the two necessary conditions of successful explanations. They are the inquiry into the *consistency* of a theory, and the inquiry into the *completeness* of a theory. Since these tools are the basis of any criticism of an equilibrium explanation, I briefly explain how they are used in economics.

> *Consistency* requires that the set of assumptions (which form any particular theory) does not lead to inconsistencies such as would be the case if both a given statement *and its denial* were logically allowed by our theory. For example, the statement 'the economy at time *t* is on its production possibilities curve' and its denial 'the economy is not on that curve' could not both follow from a consistent theory. This requirement, however, does not rule out the possibility of a theory allowing for competing or contrary situations such as multiple equilibria. For example, all points on a production possibilities curve are potential equilibria that differ only with regard to the given price ratio. If there is a flat spot on the curve, there is a set of points (along the flat spot) all of which are potential equilibria for the same price ratio.

Thus, if our explanation of why the economy is at one particular point along the flat spot is that it is faced with the corresponding price ratio, then consistency alone will not enable us to explain why the economy is not at any other allowed point on the flat spot. Nevertheless, consistency is obviously important since we cannot tolerate contradictions or inconsistencies.

> *Completeness* is the requirement that an explanation does not allow for the possibility of competing or contrary situations. Completeness rules out the possibility of a false explanation accidentally appearing to be true. That is, if our explanation is complete and happens to be false, we shall be able to show it to be false directly. For example, if we assume that the production possibilities curve has no flat spot and is concave (to the origin) then our explanation would be logically complete since each point on the curve is compatible (tangent) with only one price ratio and each price ratio is compatible with only one point on the curve. In other words, our equilibrium point is *unique* given any particular price ratio. Should any other equilibrium point be possible for the same price ratio, then we would also have to explain why we observe the one point rather than the other possible points. That is, our model must explain why we do not observe what we do not observe. The logical possibility of other compatible points would mean that our model is not complete.

The standard method of demonstrating the consistency of a theory is to construct a mathematical model of that theory and prove that it necessarily possesses a sensible solution – that is, demonstrate the *existence* of a sensible solution. The standard method of demonstrating the completeness of a theory is to show that the equilibrium solution of the model is *unique*. Although there is some danger of confusion, these two attributes of theories are usually analyzed separately. There are other, secondary, aspects of axiomatics such as inquiries into the independence and 'weakness' of the various assumptions that make up a theory. I will not discuss these topics here since they are questions of aesthetics rather than of the explanatory power of any equilibrium model.

Usually the question of consistency can be dealt with in a rather direct way: try to solve the system of equations constituting the model of the theory. If a sensible solution cannot *always* be obtained, it may be possible to specify additional assumptions to guarantee such a solution. Eliminating non-sensible solutions is a low-order completeness criterion – that is, the model must be complete enough to exclude them but it may not be complete enough to allow only one sensible solution.

The conditions which assure consistency are usually much less restrictive than those which assure completeness. For this reason the question of completeness can be a serious source of important fundamental criticism. One of the pioneers of axiomatic analysis in economics, Abraham Wald, offered such a criticism of Walrasian economics. A well known but minor aspect of his analysis was a simple proof that the popular

condition that 'the number of equations be equal to the number of unknowns' was neither necessary nor sufficient to guarantee a solution, let alone a unique solution. Wald's 1936 axiomatic study of Walrasian general competitive equilibrium, which now may be merely of interest to historians of mathematical economics, can serve as an interesting case study to demonstrate the importance of completeness. Subsequently, I will present my theory of completeness which I think is relevant for general economists as well as for mathematically oriented theoretical economists and which I think may be the only effective means of criticizing equilibrium models.

WALD'S AXIOMATIC WALRASIAN MODEL: A CASE STUDY

Rarely will we find axiomatic studies of Marshallian economics. The reason is simple but misleading. The reason is that Marshall's statical method focuses primarily on the necessary equilibrium requirements for just one market at a time. The key notion is a *partial equilibrium* which is partial because all other markets are impounded in the *ceteris paribus* condition invoked in the determination of each individual's demand (or supply). But each individual still needs to know the prices of other goods. In other words, the individual makes substitution choices on the basis of a knowledge of relative prices. Thus, in effect, the partial equilibrium method is actually predicated on all other markets providing equilibrium prices – otherwise, the equilibrium of the market in question will not persist. The absence of such a general market equilibrium will usually lead to price changes in the other markets followed by appropriate substitution responses in the demand and supply curves of the market in question. So ultimately a complete Marshallian explanation of an equilibrium price involves a form of general equilibrium since only when there is a general market equilibrium can we be sure there is a partial equilibrium in the market in question. Thus Marshall and Walras differ only in their methodological procedures. Since the ultimate equilibrium state in one market depends on all other markets being in equilibrium, the most direct way to analyze the requirements of a general market equilibrium would be to consider all individuals simultaneously and try to determine a set of prices that would allow all individuals to be maximizing. This latter procedure is the Walrasian approach to equilibrium explanations. Although Marshall's procedure may appear to differ, any analysis of a Walrasian equilibrium state will have implications for any successful application of the statical method even when focused on just one market.

The Walrasian system of general equilibrium thus purports to explain simultaneously all (relative) prices and all (absolute) quantities of traded goods (in the system). The question of interest here is: What is the logical consequence of the assertion that Walras' system does *explain* all the (endogenous) variables? In particular, what are the logical conditions placed on the system for it to be truly 'in equilibrium'? When we say the system explains all the prices and quantities, we are saying that all the explicit and implicit (i.e. unstated) assumptions necessary for the sufficiency of the explanation are satisfied. In other words, we are claiming that the system of assumptions is complete. We know what the explicit assumptions are in Walras' system, but the question remains, what are the implicit assumptions? To conjecture what the implicit assumptions are is the task of an axiomatic analysis of the completeness of a general equilibrium system such as Walras'. However, before the search for implicit assumptions can begin, we must first show that the explicit assumptions form an incomplete system, that is, an incomplete system with respect to the task of explaining all prices and quantities of traded goods. Wald, in his famous 1936 paper, attempted to do both of these tasks, namely, to demonstrate the incompleteness of the Walrasian system (which supposedly Walras at first thought was complete merely because the number of equations equalled the number of unknowns) and to posit some possible implicit assumptions. His paper represents one of the first rigorous (axiomatic) studies of the mathematical implications of a Walrasian economic system (in general equilibrium).[3] His version of a Walrasian system is the following:

$$\left.\begin{aligned} r_i &= a_{i1}X_1 + a_{i2}X_2 + \ldots + a_{in}X_n + U_i && (i=1, 2, \ldots, m)\\ U_iV_i &= 0 && (i=1, 2, \ldots, m)\\ P_j &= \sum_{i=1}^{m} a_{ij}V_i && (j=1, 2, \ldots, n)\\ P_j &= f_j(X_1, X_2, \ldots, X_n) && (j=1, 2, \ldots, n) \end{aligned}\right\} \quad [4.1]$$

where the exogenous variables are as follows:

r_i is the quantity available of the ith resource

a_{ij} is the quantity of the ith resource needed per unit of the jth good

and the endogenous variables are as follows:

U_i is the unused portion of the available ith resource

P_j is the price of the jth good

V_i is the value of the ith resource

X_j is the output quantity of the jth good

This system of equations is the beginning of an axiomatic version of a Walrasian economic system. The first class of equations (r_i = ...) represents the production or resource allocation relations. The second class is a special consideration which says that if a resource is not scarce then some of it will be unused ($U_i > 0$), and thus the resource price (V_i) must be zero (i.e. it is a free good). Walras was claimed to have ignored this consideration (perhaps because he thought it would be obvious which resources are scarce). The third class of equations is the typical long-run competitive equilibrium condition where price equals unit cost. Now the fourth class is actually a set of Marshallian market demand curves. Wald's axiomatic version of the Walrasian system then differs slightly from the textbook version of Walrasian neoclassical economics. In particular, his version makes no attempt to explain the market demand curves by explaining *individual* consumer behaviour.

Wald's study involved the question 'Does the system of equations [4.1] have a unique non-negative system of solutions where r_i and a_{ij} are given numbers, $f_i(X_1, ..., X_n)$ are given functions, and the U_i, X_i, V_i and P_i are unknowns?' On the basis of his method of rationalizing his affirmative answer to this question, he formulated the following theorem which he said he proved elsewhere [Wald 1933/34, 1934/35].

Theorem. The system of equations [4.1] possesses a set of non-negative solutions for the $2m + 2n$ unknowns and a unique solution for the unknowns $X_1, ..., X_m, P_1, ..., P_n, U_1, ..., U_m$, *if* the following six conditions are fulfilled:[4]

(1) $r_i > 0$ $(i=1, 2, ..., m)$.

(2) $a_{ij} \geq 0$ $(i=1, 2, ..., m; j=1, 2, ..., n)$.

(3) For each j there is at least one i such that $a_{ij} > 0$.

(4) The function $f_j(X_1, X_2, ..., X_n)$ is *non-negative* and *continuous* for all n-tuples of non-negative numbers $X_1, X_2, ..., X_n$ for which $X_j \neq 0$ $(j=1, 2, ..., n)$.

(5) If the n-tuple of non-negative numbers $X_1^k, ..., X_n^k$ $(k=1, 2, ... \infty)$ in which $X_j^k > 0$ for each k, converge to an n-tuple $X_1, ..., X_n$ in which $X_j = 0$, then

$$\lim_{k\to\infty} f_j(X_1^k, X_2^k, ..., X_n^k) = \infty \quad (j=1, 2, ..., n).$$

(6) If $\Delta X_1, \Delta X_2, ..., \Delta X_n$ are *any n* numbers in which *at least one* < 0, and if

$$\sum^n P_j \Delta X_j \leq 0,$$

then

$$\sum^n P_j' \Delta X_j < 0,$$

where $P_j' = f_j(X_1+\Delta X_1, ..., X_n+\Delta X_n)$ $(j=1, 2, ..., n)$.

Furthermore, he noted that if the rank of the matrix $[a_{ij}]$ is equal to m, then the solution is also unique for the variables $V_1, ..., V_m$.

Now let us try to see what Wald has imposed on the well known Walrasian economic explanation of prices and outputs. The first three conditions are the usual economic considerations. Condition (1) says that the resources must exist in positive amounts in order to be used. Condition (2) says that input requirements are not negative (i.e. they are not outputs). And condition (3) says the output of any good must require a positive amount of at least one input.

Conditions (4) and (5) are required for the *method* of proving *his* existence and uniqueness theorem. That is, in order to use calculus-based mathematics, he must simplify the mathematical aspects of the system. But, whereas condition (4) involves only the usual assumption of continuity, condition (5) is a more serious simplification. Condition (5) says that for the quantity demanded of a good to be zero, the price must be infinitely large. He says that this condition is not necessary for an existence proof but it does help by making the mathematics simple (this condition was the first to be discarded by subsequent developments in mathematical economics twenty years later).[5]

Now we reach (6), the most important condition. It is so important that it has been given a special name: the Axiom of Revealed Preference.[6] It says that the demand functions must be such that if combination A of goods is purchased rather than any other combination B that cost no more than A at the given prices then, for combination B ever to be purchased, the prices must change such that combination B costs less than combination A at the new prices. A rather reasonable assumption if we were speaking of individual consumers, but these are market demand curves! Unfortunately, it does not follow that if the axiom holds for each individual consumer's demand function, then it necessarily will hold for the market function. Similarly, when it holds for the market, it does not necessarily hold for all the individuals. One behavioural interpretation of condition (6) is that all consumers act alike and thus are effectively one. Thus condition (6) imposes constraints on the 'community indifference map' which may be difficult or impossible to satisfy.

We should thus ask (as did Wald): Do we *need* the axiom of revealed preference (in order to assure completion)? His answer was 'yes', and he demonstrated it with a simple model of system [4.1]. Note that if it is necessary for system [4.1] it is necessary for every model of the system;

thus if we could show that it is unnecessary for any one model, we could refute its alleged necessity for the systems.

Conditions (1) to (5) are necessary for Wald's proof of the *consistency* of his version of the Walrasian system. Condition (6) is necessary to *complete* the system. To show this we shall specify a model which satisfies conditions (1) to (5), and then we show the necessity of condition (6) by describing a case in which condition (6) is not fulfilled and for which a unique solution does not exist. Consider Wald's special case of system [4.1] involving the unknowns X_1, X_2, P_1, P_2 and V_1:

$$\left.\begin{aligned} r_1 &= a_1X_1 + a_2X_2 \\ P_1 &= a_1V_1 \\ P_2 &= a_2V_1 \\ P_1 &= f_1(X_1, X_2) \\ P_2 &= f_2(X_1, X_2) \end{aligned}\right\} \qquad [4.1']$$

And to satisfy conditions (1), (2) and (3), we can simply let $a_1 = a_2 = a$ where $a > 0$ and let $r_1 > 0$. To satisfy (4) we assume $f_j(X_1, X_2)$ to be continuous and positive. To satisfy (5) we assume that as X_j approaches zero, $P_j \to \infty$. The heart of the matter is the inverse demand functions, $f_j(X_1, X_2)$.

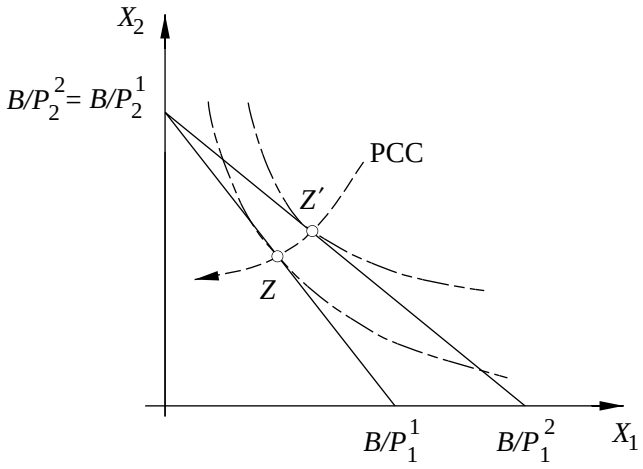

Figure 4.1 *Price–consumption curve (PCC)*

Let us therefore look more closely at them by first reviewing textbook indifference analysis, and in particular, we want to look at the nature of the set of combinations of X_1 and X_2 which give the same demand price (i.e. for P_j constant).

We know that whenever we base consumer theory on indifference analysis we can derive the demand curve for a good by considering what is usually called the 'price–consumption curve'. To illustrate, consider the two goods, X_1 and X_2. Specifically, all the possible non-negative combinations of them, and let us assume that income is given. Note that in Figure 4.1, for a particular combination of goods, say point Z, there is only *one* set of prices which will be compatible with a choice of combination Z, in particular $P_1{}^1$ and $P_2{}^1$. If we were to change $P_1{}^1$ to $P_1{}^2$ without changing P_2, we should find that point Z' is the combination which is compatible with the new price(s).[7]

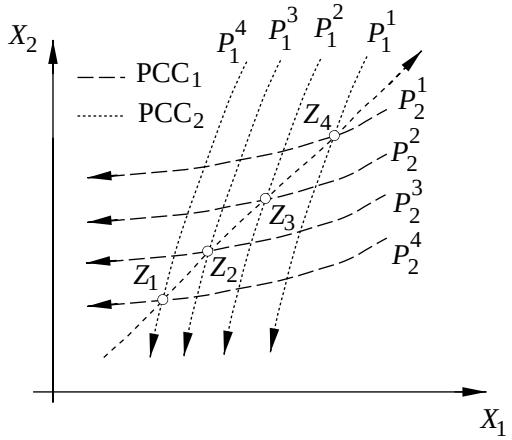

Figure 4.2 *The Z-line (income–consumption curve)*

In this manner we can trace all the combinations which are compatible with a particular P_2 (i.e. where P_2 is constant). The curve traced is simply the price–consumption curve for X_1 from which we derive the demand curve for X_1 or, in terms of model [4.1′], it is all the combinations of X_1 and X_2 such that $f_2(X_1, X_2) = \text{constant}$. Now, instead of drawing an indifference map, we could simply draw a representative set of the possible price–consumption curves (assuming income given) and get something like Figure 4.2. In this figure each curve is labelled with the appropriate fixed level representing the fixed price of the *other* good. On this diagram we can see that point Z_1 is compatible only with given prices $P_1{}^4$ and $P_2{}^4$. If we hold P_2 constant and move outward from point Z_1, in neoclassical consumer theory we should find that P_1 *falls* along the price–consumption curve labelled with the fixed price $P_2{}^4$ (see also Figure 4.1). Similarly, if we hold P_1 constant and move outward along the other price–consumption curve from Z_1, then P_2 falls. Thus note in Figure 4.2 that the superscripts

indicate an ordering on prices. Also we note that conditions (4) and (5) can be satisfied; for example, as we move horizontally toward the vertical axis (i.e. X_1 goes to zero) the price of X_1 rises. If we let $P_1{}^1 = P_2{}^1$, $P_1{}^2 = P_2{}^2$, ..., $P_1{}^k = P_2{}^k$, we can trace all the combinations for which $P_1 = P_2$, viz. Z_1, Z_2, Z_3, etc. The line connecting these Zs is what is usually called the 'income–consumption curve' but since the definition of price–consumption curves is based on a fixed budget or income, I will call this the Z-line.

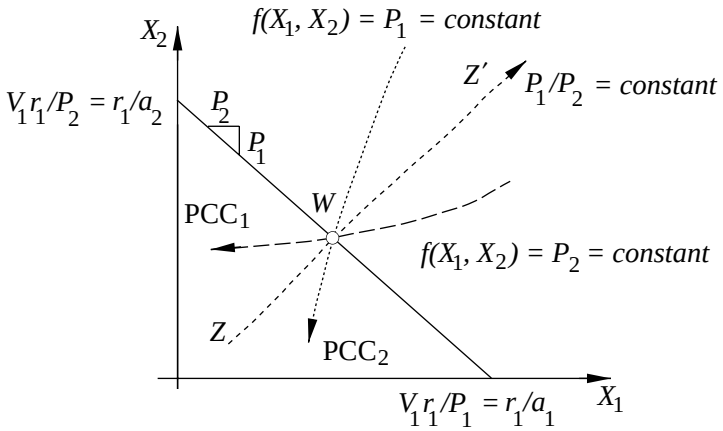

Figure 4.3 *Price–consumption curves and Wald's special case*

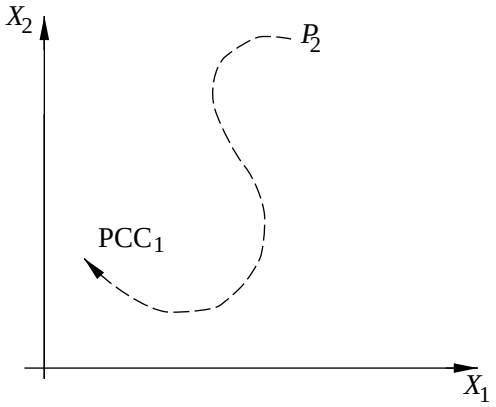

Figure 4.4 *A denial of condition (6)*

Returning to system [4.1′], we see that the first equation can be represented on the commodity–space diagram as shown in Figure 4.3. Since r_1, a_1 and a_2 are given we describe the set of combinations of X_1 and X_2 which satisfy the first equation as a line (resembling a budget line) which satisfies conditions (1), (2) and (3). Condition (4) says that through each and every point in Figure 4.3 there is exactly one price–consumption curve for good X_1 and exactly one for good X_2. Condition (5) says that as we trace out any price–consumption curve for good X_1 in the direction indicated by the arrowhead (i.e. for a rising P_1) the price–consumption curve will never touch the X_2 axis. Condition (6) is less obvious. It says that no price–consumption curve for good X_1 will have a shape illustrated in Figure 4.4.[8] The reason for excluding such a shape is that the inverse demand function implied by such a shape might not be sufficiently well defined. Condition (6) also assures a sufficient degree of convexity of the underlying preference map (which would have to be a community's map in Wald's model). In my diagrams, this means that if you face in the direction indicated by the arrowhead on any particular Z-line, then to your left the ratio of P_1/P_2 will always be higher than the one corresponding to this Z-line.

What Wald's proof establishes is that there is at least one stable equilibrium point on the quasi-budget line through which passes the correct Z-line. The correct Z-line will be the one drawn for a P_1/P_2 ratio that equals the slope of the quasi-budget line. That is, he proves that there is at least one point like either the one on the positively sloped Z-line illustrated in Figure 4.3 or like the one on a negatively sloped Z-line which has its arrowhead outside of the feasible production points limited by quasi-budget line as illustrated in Figure 4.5.[9]

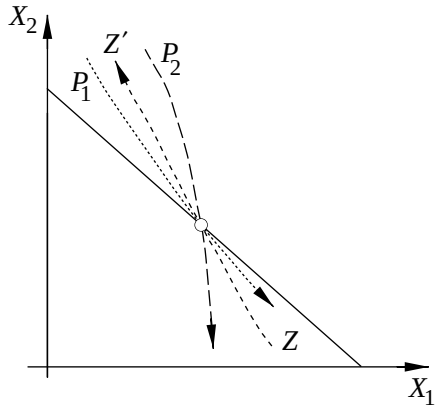

Figure 4.5 *A possible negatively sloped Z-line*

COMPLETENESS AND THEORETICAL CRITICISM

Although the inclusion of Wald's six conditions in the axiomatic structure of the Walrasian system fulfills the task of completing an explanation of prices and outputs, it does not follow that they are *necessary* for the *original* theory. As it was later shown, the existence and uniqueness of the entire Walrasian system can be proved by using either linear programming or activity analysis and these do not require such restrictive assumptions. Thus it would seem that if we are able to show that any one of Wald's conditions is not satisfied (in the 'real world') we do not necessarily refute the original incomplete theory. From a methodological position, this state of affairs is rather perplexing. We may wish to complete an axiomatic version of neoclassical price theory and then criticize it. But, if our criticism deals only with those conditions which we add (for completion purposes), then we are not really criticizing the original price theory. Some think this can be overcome by attempting to deduce testable statements from the incomplete theory and submitting these to tests. No matter how the theory is eventually completed, should any one of them be shown to be false, the theory *as a whole* will be false – otherwise, the apparent falsifying fact must be explained away! Either way, this is a very difficult task and not much has been attempted or accomplished so far.[10]

The question of testability (or criticizability in general) is above all a logical problem. And since axiomatic analysis is concerned with the logical properties of a theory, it can have something to say about empirical testability as well as being able to offer a means of theoretically testing a theory. For example, we should probably view most of the theoretical analysis of neoclassical textbooks as failures of indirect attempts to test the completeness of the neoclassical theory (i.e. failures to show the neoclassical theory to be incomplete). Actually, what we read in the textbooks should be viewed as the only aspects of the theory which are considered complete (often only on the basis of apparent, but untested, consistency).

This disagreement in viewpoints is not just apparent. It would seem that few economists are directly concerned with completeness because most of them (implicitly or explicitly) view economic knowledge as a logical system which is supported by positive evidence. 'Supported' usually means that at least some predictions (or propositions) that logically follow from their theories have been verified or confirmed. An unintended outcome of this view of knowledge is that most economists are satisfied with an argument whenever it allows for the *possibility* of the truth of their theory even though the theory at the same time may imply propositions which are false. For example, a model may have several solutions, one of which is true (i.e. agrees with the observed facts) but the others are false. A completed model, however, leaves no room for errors (viz. for disagreement with facts). Unfortunately, most economists would be satisfied with the incomplete model because at least one of its many solutions is true.

There are different theories of knowledge. Obviously, the one I am promoting in this book says the only way we learn is through criticism; and of course, testing is one form of criticism. Incomplete theories are very difficult to criticize because they leave so much room for conceivable contradictions. Because I want to learn, I want to be able to criticize any theory, and attempting to complete a theory is an important means of exposing a theory to decisive criticism. The unintended outcome of this view of knowledge is that when we attempt to explain an economic equilibrium (such as Walras') it is necessary to explain why all other possible equilibrium positions are not obtained. In effect, this says we must be concerned with *uniqueness*, since to be complete (and thus testable) our explanation of any alleged equilibrium must not allow for other contrary situations such as 'multiple equilibria'.[11] This view is contrary to the popular myth (all too often promoted by those economists who 'picked up mathematics on the side') that satisfying the calculus conditions of a 'stable equilibrium' is sufficient to explain the equilibrium in question. A stable equilibrium structure (such as a negatively sloped demand curve and positively sloped supply curve) is necessary, of course, but without behavioural assumptions concerning price adjustment dynamics, we still have not explained why the system is in 'equilibrium' where it is. All that the calculus stability conditions accomplish is the avoidance of confusing a possibly unstable 'balance' situation with a stable equilibrium situation. I will return to the matter of the importance of stability conditions in Chapter 14.

A THEORY OF COMPLETENESS

In spite of what economists think they are doing, they can be seen to have been indirectly concerned with completeness, and the evidence is the development of neoclassical economic theory. One way to understand this development on the basis of a theory of the development of theories is to characterize all theories as systems of assumptions where each assumption is in the logical form of an 'all-and-some' statement. As I briefly discussed in Chapter 1, an 'all-and-some' statement is one of the form 'for all x there is some y such that ...'. The 'such that ...' clause may or may not be completely specified depending on whether or not, and to what extent, the theory has been completed. Thus an attempt, such as Wald's, to complete a

model of a Walrasian theory is in effect an attempt to specify the 'such that ...' clauses of the theory. Whether an 'all-and-some' statement is empirically testable is a question of *how* the 'such that ...' clause has been completed. It is always possible to complete a theory without making it testable; for example, by making it circular.[12]

The specification of the 'such that ...' clauses is almost always *ad hoc*, and so is the completion of an axiomatic system. The history of formal model-building in neoclassical economics is one of a sequence of efforts to complete systems of ideas which rationalize certain enduring propositions. The specification of the nature of indifference curves by Hicks and Allen [1934], the specification of imperfect competition by Robinson [1933/69], the specification of the idea of a market equilibrium by Samuelson [1947/65], and the attempts of Franco Modigliani [1944] and Donald Patinkin [1956] to explain Keynes, are all examples of developments in the neoclassical theory which amount to completions of 'such that ...' clauses. These are also examples of placing requirements on theories which are similar to requirements of typical axiomatic analyses.

If an axiomatic analysis of a theory manages to posit requirements which are necessary for the sufficiency of any given model of that theory, it is an important achievement which should not be left only to mathematical economists to pursue. Wald's Axiom of Revealed Preference, for example, is such a requirement. Any requirement (or 'condition') that is necessary for the completion of a theory may offer an important opportunity for critically testing that theory. However, the Axiom of Revealed Preference by itself is not an essential element in economic analysis.[13] What is essential in neoclassical economics is the notion of a state of equilibrium. In the next chapter I examine other ways to view equilibrium analysis.

NOTES

1 Of course, there are some neoclassical economists who even put the existence of a state of equilibrium beyond question.

2 This type of analysis began in the nineteenth century with studies of the axiomatic structure of Euclid's geometry [see Blanché 1965].

3 Many other axiomatic studies have been published since Wald's, for example Arrow and Debreu [1954], Arrow and Hahn [1971], Debreu [1959, 1962], Gale [1955], McKenzie [1954, 1959].

4 Note well that he does not say *only if*.

5 Specifically, by replacing it with a duality assumption [see Kuhn 1956]. It should be noted that Wald recognized the possibilities of using other mathematical techniques which did not require such a condition. See Quirk and Saposnik [1968] for a survey of the other well-known axiomatic studies of Walrasian economics.

6 Today, Wald's condition is called the *Weak* Axiom of Revealed Preference since it is limited to the comparison of two points. A strong version would refer to a chain of comparisons of many points [see Houthakker 1950, 1961]. None of the discussion in this book will require us to be concerned with this distinction so I will not be emphasizing the 'weakness' of this axiom.

7 The arrowhead on the price–consumption curve indicates the direction along which the changing price increases for the given income and price of the other good.

8 This interpretation of the Axiom of Revealed Preference will be the subject of Chapter 13.

9 Note Figure 4.5 can be used to represent two kinds of appropriate *Z*-lines simply by swapping the X_1 and X_2 labels (and the P_1 and P_2 labels).

10 Paul Samuelson has in effect attempted to deal with this in Chapter 5 of his published PhD thesis [1947/65]. I have discussed his attempt in Boland [1989, Chapter 1].

11 Whether multiple equilibria represent contrary situations depends on what we are trying to explain. For example, if we were trying to explain the price–quantity in market *A* and we found that it was compatible with various equilibria in market *B*, there would be no problem. But, if there are various possible equilibria in market *A* allowed by our explanation of market *A*, then we have an incomplete explanation.

12 To the statement 'for every rationalizable choice there is a maximizing choice ...' we might add 'such that if it is not a maximizing choice it is not rationalizable'.

13 The axiom just happens to be the one used in Wald's and others' attempts to formally analyze their invented models of neoclassical equilibrium. The *role* of this axiom in the formalization of neoclassical economics will be further explored in Chapter 13.

5 Axiomatic analysis of disequilibrium states

> The theory of stable equilibrium of normal demand and supply helps indeed to give definiteness to our ideas; and in its elementary stages it does not diverge from the actual facts of life, so far as to prevent its giving a fairly trustworthy picture of the chief methods of action of the strongest and most persistent group of economic forces. But when pushed to its more remote and intricate logical consequences, it slips away from the conditions of real life.
>
> Frank Hahn [1973, p. 1]

While the axiomatic analysis of equilibrium models can determine whether a given model is consistent and complete, little analysis has been done concerning consistency and completeness of models of disequilibrium states.[1] Obviously, we cannot expect to be able to assess solvability as a means of assuring consistency since, as discussed in Chapter 4, the solutions of the equilibrium models were sets of equilibrium prices that could be used possibly to explain existing prices. In this chapter I will offer a few elementary axiomatic analyses of models of 'disequilibrium' states. Eventually, we will need to consider how they may be used to critically assess any axiomatic analysis of disequilibrium models.

There are two ways to use disequilibrium models. One is to *explain why* disequilibrium phenomena occur and the other is to *explain away* disequilibrium phenomena as mere appearances. Both utilize underlying equilibrium models in which it is assumed that all consumers are maximizing utility (either directly or indirectly by maximizing personal wealth) subject to given equilibrium prices and all producers are maximizing their profit subject to given technology and given market equilibrium prices.

Since virtually all neoclassical equilibrium models take for granted that there are no barriers to any consumer quickly responding to changing prices, if there is a state of disequilibrium, such a state will be found by analyzing the logic of the situation facing the producers.[2] In this chapter, I will follow this tradition by focusing on the theory of the individual producer to determine how the logic of the situation facing the firm may be used to account for any state of disequilibrium.

COMPETITION BETWEEN THE SHORT AND LONG RUNS

In regard to the theory of the firm facing a general equilibrium situation, I want to examine the role played by two particular assumptions. One is the assumption that prices are *fixed givens* which in turn is based on an assumption that the firm is a 'perfect competitor' (perhaps because it is too small to be able to affect its price by altering the supply). I wish to show why dropping the fixed-price assumption would severely restrict our choice of assumptions regarding other aspects of the firm. The other assumption to be examined is one concerning the applicability of the assumption of profit maximization. In Chapter 3 I noted that Marshall defined a short run where everything but the input of labour and the level of resulting output are fixed. At the other extreme is his long run where everything but technology is variable (and thus subject to his Principle of Substitution). Here I will examine what might transpire in the shadowy area between Marshall's short and long runs, that is, in what I will call the *intermediate run*. The distinction between the Marshallian runs is solely a matter of the time available in the period under consideration and a recognition that some inputs are easier to change than others (i.e. change takes less time). In Marshallian terms (i.e. assuming just two inputs, labour and capital[3]) the question is the speed by which capital can be physically changed. While it is commonplace to define the short run as a period of time so short that there is not enough time to change capital, the long run presumes that both inputs are unrestrictedly variable. Now, the purpose of recognizing an intermediate run is to recognize that there are two ways of changing capital, internally and externally. The period of time corresponding to the intermediate run is defined to be too short to allow wholesale changes in the physical type of the capital used in the firm but long enough to allow the firm to vary internally the quantity of the existing type of capital used. In the intermediate run the firm must decide upon the *optimum quantity* of capital. In the long run, however, there is sufficient time to change to a different type of capital as is usually the case when a firm switches from one industry to another. Thus, in the long run the firm must decide upon the *optimum type* of real capital.

One reason why many theorists wish to drop either the perfect-competitor assumption or the profit-maximizer assumption is simply that these assumptions in many cases are 'unrealistic' in disequilibrium models.

Some just complain that these assumptions are plainly 'unrealistic' in the sense that it would be realistic to assume that the firm is a perfect competitor only when there are an extremely large number of firms, each of which is relatively small – for example, an economy of 'yeoman farmers' or perhaps an economy consisting of only small businesses. A small firm has to take its product's price as given only because it will go out of business if a higher price is charged since its customers can go to any of the large number of competing firms. Similarly, if it charges less than the given price when the given price is the 'long-run equilibrium price' (which equates with average cost) then it will be losing money and will still eventually go out of business. It is thus said that with a large number of small firms competition can be 'perfect'.

Would-be 'realists' argue that the modern economy consists of relatively large firms or few firms in each industry (or both) and thus, they say, in the real world there is 'imperfect' competition. Imperfect competition allows for two possible circumstances. First, it is possible for the firm to be a price-taking 'competitor' and also be one of a few producers such that changes in its output do affect the *market*-determined equilibrium price. The second is to assume that the firm is a price setter such as the usual textbook's monopolist. The first approach will be the one adopted here since it does not require the producer to know the full nature of the demand curve facing the firm. The second approach can be considered a special case of the first – namely where the firm's demand curve is the market's demand curve *and* the firm has full knowledge of the market.

THE 'PERFECT-COMPETITOR' FIRM IN THE LONG RUN: A REVIEW

In order to examine the axiomatic role of the assumptions of the Marshallian theory of the firm, we need to discuss the effect that dropping the perfect-competitor assumption would have on equilibrium models and in particular on the assumptions concerning the production function. Before we drop this assumption, however, let us review the basic logic of the perfect-competitor firm with respect to its production function.

Since by definition the intermediate run involves less time than the long run, it can be argued that a long-run equilibrium must also be an intermediate-run equilibrium and similarly it must also be a short-run equilibrium. Most important in the recognition of the intermediate run is the separation of the zero total profit idea ($TP = 0$) from the idea of *complete* profit maximization (i.e. with respect to all inputs). To do this we need to recognize the explicit conditions necessary for each of the three types of equilibria. In the short run, since only labour can be varied, an equilibrium is reached once the optimum amount of labour has been hired. The necessary condition for this is that the price of the good being produced equals its marginal cost (MC) or, in terms of the decision concerning labour, that the marginal physical product of labour (MPP_L) equals the real cost of one unit of labour. Specifically, the existence of a *short-run equilibrium* assures us that $MC = P_x$ or $MPP_L = W/P_x$ (where the good produced is X and the prices of X and labour are, respectively, P_x and W). Given a price of capital (P_k), an *intermediate-run equilibrium* assures that the optimum quantity of capital has been utilized such that the marginal product of capital (MPP_K) equals the real cost of capital (P_k/P_x). And since the intermediate run is longer than the short run (i.e. there is sufficient time to satisfy both sets of conditions), we can also be assured that the marginal rate of technical substitution ($MRTS$) between labour and capital equals the relative costs of those inputs (W/P_k). Except when we limit the notion of a production function to the special case of linear-homogeneous production functions, we will see that the attainment of an intermediate-run equilibrium does not assure a long-run equilibrium. Specifically, an intermediate-run equilibrium will not assure us that total profit is zero. The absence of zero total profit means that there may be an incentive for new entries or exits and thereby means that there may be incentives which deny an equilibrium state (since there is sufficient time for such reactions).

Most textbooks go straight to the long-run equilibrium from the short-run equilibrium. That is, they go from where, while $MPP_L = W/P_x$, it is possible that $MRTS \neq W/P_k$ (since not all short-run equilibria are long-run equilibria) to a long-run equilibrium where $MRTS = W/P_k$ and $TP = 0$. It is interesting to note that the long-run equilibrium is the starting point for an Adam Smith type of philosophical discussion of the virtues of competition and self-interest. That is, if every firm is making 'zero profits' with the given production functions (i.e. given technology) the only way a firm can obtain positive 'excess' profits is to develop new cost-reducing technologies. In the absence of competition such 'greed' (in this case, the pursuit of extra profits) would mean that one firm might gain at the expense of others, but if we also have 'free enterprise competition' any improvements in productive efficiency which reduce costs will eventually be shared by all the firms and thus benefit everyone through lowered prices.

All this seems to be taken for granted or ignored in most textbooks. Everyone seems to be satisfied with discussing only the necessary properties of the *long-run* equilibrium – as if there were virtue in zero profit itself! There is some virtue to having the lowest possible price for a given technology but it leaves open the question from a broader perspective of the choice of optimal production or the optimal 'quality' of capital and its associated technology.

What the recognition of an intermediate-run equilibrium allows is the discussion of situations where profit is maximized with respect to all inputs but $TP \neq 0$. The basis for this discussion is that while zero profit is due to decisions which are external to the firm, the efficiency of production ($MRTS = W/P_k$) is due to an internal decision whereby profit is maximized with respect to *all* inputs. The intermediate run is often ignored because the properties of the long-run equilibrium are considered more interesting – usually, this is because they are mathematically determinant and thus available for applications. Unfortunately, the long-run equilibrium conditions are considered so interesting that models of the firm are designed to guarantee that it is logically *impossible* to have an intermediate-run equilibrium which is not a long-run equilibrium. I shall now show how this is done and as well show how such models are also incompatible with imperfect competition.

PROFIT MAXIMIZATION WITH CONSTANT RETURNS TO SCALE

The basic ingredient of long-run models of the firm is the assumption that the production function is 'linear-homogeneous' (e.g. doubling all inputs will exactly double output) – this is usually called 'constant returns to scale'. As stated, this assumption is *not* a necessary assumption for the attainment of a long-run equilibrium since the existence of such an equilibrium only requires the existence of a point on the production function which is *locally* linear-homogeneous [see again Baumol 1977, p. 578]. However, it is not uncommon for a long-run model-builder to assume that the production function is *everywhere* linear-homogeneous.

Parenthetically, let us note that a production function will necessarily be linear-homogeneous if *all* inputs are unrestrictedly variable.[4] But, if any input is fixed (such as space, time available, technological knowledge, management talents, etc.) or cannot be duplicated, then the relationship between the other inputs and the output will not usually be everywhere linear-homogeneous.

For now, let us examine the properties of everywhere-linear-homogeneous production functions. First let us note that the homogeneity of such a function implies Euler's theorem holds, that is, for any function $X = f(L,K)$ it will be true that:

$$X = MPP_L \cdot L + MPP_K \cdot K \quad \text{at all } L, K \text{ and } X = f(L,K). \qquad [5.1]$$

Now I shall show that when one adds to this assumption that the firm is in an intermediate-run equilibrium one automatically obtains the necessary conditions for a long-run equilibrium. The intermediate-run equilibrium assures that, *given* P_x (as well as W and P_k), whenever the firm is internally maximizing profit with respect to both labour and capital, the following two equations are true:

$$MPP_L = W/P_x \qquad [5.2a]$$

$$MPP_K = P_k/P_x. \qquad [5.2b]$$

Now, the combination of [5.1], [5.2a] and [5.2b] leads to the following:

$$X = (W/P_x)\cdot L + (P_k/P_x)\cdot K \qquad [5.3]$$

or rearranged by multiplying both sides by P_x:

$$P_x \cdot X = W \cdot L + P_k \cdot K. \qquad [5.3']$$

The left side of [5.3′] is total revenue (TR) and the right side is total cost (TC), hence it implies $TP = 0$. This means that in the usual long-run model, with its typical *everywhere*-linear-homogeneous production function, intermediate-run equilibrium implies all necessary conditions of long-run equilibrium. That is to say, one cannot obtain an intermediate-run equilibrium *without* obtaining the necessary conditions for a long-run equilibrium of the firm.

Given that we try to explain to students the importance of competition for the attainment of a social optimum (i.e. an efficient allocation of society's resources that allows for all parties to be maximizing), it is curious that many model-builders so glibly assume the existence of constant returns to scale. If competition is to matter, the production function *cannot* be everywhere linear-homogeneous. It is the *external* pressure of competition that eventually produces the condition of zero profit (if profits are positive there is an incentive for someone to enter the competition from outside the industry).

At this stage of the discussion,[5] an important general limitation regarding assumptions [5.1], [5.2a], [5.2b] and [5.3] should also be noted. Specifically, *whenever any three of the statements are true, the fourth must also be true*. For example, this means that even when it is impossible to vary the amount of capital used and yet the production function is everywhere linear-homogeneous, if there is enough time for a short-run equilibrium and for competition to force profits down to zero, the firm will unintentionally be maximizing profit with respect to its fixed capital.[6] Similarly, even if there is no reason for the production function to be everywhere linear-homogeneous, maximization and competition will force the firm to operate at a point where the production function is at least locally linear-homogeneous.

LINEAR HOMOGENEITY WITHOUT PERFECT COMPETITION

Note that what is accomplished with the assumption that the firm is a perfect competitor is to allow P_x to be used as it is in [5.2a]. That is, if P_x is given, P_x is both average revenue (*AR*) and marginal revenue (*MR*). Thus, [5.2a] can be rearranged according to the definition of marginal cost (*MC*)[7] to obtain:

$$P_x = MC. \qquad [5.2c]$$

Equation [5.2c] is merely a special case of the more general necessary condition of profit maximization:

$$MR = MC. \qquad [5.2c']$$

Now whenever the firm is not a perfect competitor and instead faces a demand *curve* for its product rather than just a *single* demand price, [5.2c′] is the operative rule for profit maximization. Facing a (positive-valued) *downward* sloping demand curve means that the price will not equal marginal revenue – the price will only indicate average revenue. And further, the downward slope means that average revenue is falling with rising quantity and thus at all prices

$$MR < AR \equiv P_x.$$

Given the value of the elasticity of demand relative to price changes, ε, and given a specific point on the curve with that elasticity, we can calculate the marginal revenue as

$$MR \equiv AR \cdot [1 + (1/\varepsilon)]$$

which follows from the definition of the terms.[8] If we take into account that price always equals *AR* and that for profit maximization *MC* = *MR* and we recognize that a firm's not being a perfect competitor in its product market does not preclude that market from setting the output price,[9] then we can determine the relationship between price and marginal cost:

$$P_x = MC / [1 + (1/\varepsilon)]. \qquad [5.2c'']$$

And if the firm is still a perfect competitor with respect to input prices[10] then the idea expressed by [5.2a] still holds and thus the necessary conditions for profit maximization with respect to both inputs are now:

$$MPP_L = (W/P_x) / [1 + (1/\varepsilon)] \qquad [5.2a']$$

$$MPP_K = (P_k/P_x) / [1 + (1/\varepsilon)]. \qquad [5.2b']$$

Next I want to show how these last two equations affect our assumptions regarding the production function. Recall that if the production function of the firm is linear-homogeneous, then [5.1] holds, that is,

$$X = MPP_L \cdot L + MPP_K \cdot K.$$

If we assume the imperfect competitor has a linear-homogeneous production function, whenever we apply the conditions of profit maximization in the intermediate run to this, namely [5.2a′] and [5.2b′], we get:

$$X = \frac{(W/P_x)\cdot L}{1 + (1/\varepsilon)} + \frac{(P_k/P_x)\cdot K}{1 + (1/\varepsilon)}$$

or rearranging,

$$P_x \cdot X \cdot [1 + (1/\varepsilon)] = W \cdot L + P_k \cdot K$$

or further,

$$P_x \cdot X = (W \cdot L + P_k \cdot K) - (P_x \cdot X/\varepsilon).$$

Since $-\infty < \varepsilon < 0$ (because the demand curve is negatively sloped) we can conclude that whenever *MR* is positive (i.e. $\varepsilon < -1$) it must be true that:

$$P_x > (W \cdot L + P_k \cdot K)/X \equiv AC$$

or in other words there will be an excess profit of

$$TP = -(P_x \cdot X/\varepsilon) > 0.$$

Thus we can say that if the firm is not a perfect competitor but is a profit maximizer with respect to all inputs (as well as facing a linear-homogeneous production function), then total profit will be positive – that is, a long-run equilibrium is impossible.[11]

POSSIBLE ALTERNATIVE MODELS OF THE FIRM

Now let us look at all this from a more general viewpoint by recognizing the four separate propositions that have been considered.

[A] The production function is *everywhere* linear-homogeneous (i.e. [5.1]).

[B] Total profit is maximized with respect to all inputs (i.e. [5.2a′] and [5.2b′]).

[C] Total profit is zero (*TP* = 0).

[D] The firm's demand curve is negatively sloped ($-\infty < \varepsilon < 0$).

We just saw at the end of the last section that a conjunction of all four of these is a contradiction – that is, if [A], [D] and [B] are true then necessarily [C] is false. We also saw before that if [A] and [B] hold, [C] also holds if [D] does *not* hold (i.e. when the price is given).

In fact, more can be said. *When any three of these propositions are true the fourth must be false.* To see this let us first note that the traditional discussion of imperfect competition with a few large firms usually considers a long-run equilibrium where total profit is forced to zero (by competition from new firms or competing industries producing close substitutes). With these traditional models, then, [C] will eventually hold. But it is usually

also assumed that the firms are all profit maximizers ([B] holds) even when facing a downward sloping demand curve (i.e. even when [D] holds). All this implies that [A] does not hold, that is, the production function cannot be *everywhere* linear-homogeneous. Specifically, the firm must be at a point where there are *increasing returns to scale*.

So far I have only discussed the properties of *everywhere*-linear-homogeneous production functions. To see what it means to imply *increasing* returns to scale, let us now examine a production function which is homogeneous but not linear. If a production function is homogeneous, it is of a form that whenever the inputs are multiplied by some arbitrarily positive factor λ (i.e. we move outward along a ray through the origin of an iso-quant map), the output level will increase by some multiple of the same λ or, more generally, for $X = f(L,K)$:

[H] $$\lambda^n \cdot X = f(\lambda \cdot L, \lambda \cdot K).$$

Note that a linear-homogeneous function is then just a special case, namely where $n = 1$. When $n > 1$ the function gives *increasing returns* to outward movements along the scale line since the multiple λ^n is greater than λ. Note also that this is just one example of increasing returns – increasing returns do not require homogeneity. Nevertheless, it is often convenient to assume that the production function is homogeneous because the question of whether returns are increasing or decreasing can be reduced to the value of the single parameter n. Moreover, in this case, we can use the particular property of any continuous function that allows us to calculate the changes in output as linear combinations of the changes in inputs weighted by their respective marginal productivities. By recognizing that at any point on any continuous function it is also true that:

[E] $$dX = MPP_L \cdot dL + MPP_K \cdot dK.$$

If we also assume [H] holds, then if using [E] we set $dL = \lambda \cdot L$ and $dK = \lambda \cdot K$, it follows that

$$dX = \lambda^n \cdot X,$$

or in a rearranged equation form:

$$\lambda^{n-1} \cdot X = MPP_L \cdot L + MPP_K \cdot K. \qquad [5.1']$$

We see here again that equation [5.1] is the special case of [5.1′] where $n = 1$.

I now wish to put [5.1′] into a form which will be easier to compare with some later results and to do so I want to express λ^{n-1} differently. Since we really are only interested in the extent to which λ^{n-1} exceeds 1, let us calculate this directly. There are many ways to do this but let us calculate the fraction, $1/\beta$, which represents the portion of the multiple λ^{n-1} that exceeds 1, that is, let

$$\lambda^{n-1} - 1 \equiv (1/\beta) \cdot \lambda^{n-1}.$$

For later reference, note that β can be considered a 'measure' of the *closeness* to constant returns (i.e. to linearity). The greater the degree of increasing returns, the smaller will be β.

The reason why I have chosen this peculiar way of expressing λ^{n-1} will be more apparent a little later, but for further reference let me re-express [5.1′] using β rather than λ:

$$X / [1 - (1/\beta)] = MPP_L \cdot L + MPP_K \cdot K. \qquad [5.1'']$$

Let us put these considerations aside for now except to remember that a production function which gives increasing returns to scale will be expressed with $0 < \beta < \infty$ or equivalently with $(1/\beta) > 0$. A few paragraphs ago it was said that [A] is denied whenever we add [C] to [D] and [B]. Let us consider the more general case where all that we know is that [D] and [B] hold – that is, the profit-maximizing firm is facing a downward sloping demand curve in an intermediate-run equilibrium situation. First let us calculate its total cost (TC):

$$TC \equiv W \cdot L + P_k \cdot K.$$

Assuming [D] and [B] hold allows us to use [5.2a′] and [5.2b′] to get

$$TC = P_x \cdot [1 + (1/\varepsilon)] \cdot (MPP_L \cdot L + MPP_K \cdot K).$$

Now we can add [C]. Since total revenue is merely $P_x \cdot X$, zero profit means that

$$X = [1 + (1/\varepsilon)] \cdot (MPP_L \cdot L + MPP_K \cdot K)$$

or more conveniently,

$$X / [1 + (1/\varepsilon)] = MPP_L \cdot L + MPP_K \cdot K. \qquad [5.4]$$

Now we can make the comparison which reveals an interesting relationship between imperfect competition and increasing returns. First note that equations [5.1″] and [5.4] have the same right hand side thus their left hand sides must be the same as well. Thus whenever [B], [C] and [D] hold, we can say that

$$1 - (1/\beta) = 1 + (1/\varepsilon)$$

or more directly,

$$\beta = -\varepsilon \; ! \qquad [5.5]$$

While we have obtained [5.5] by assuming that the imperfect competitor is in a long-run equilibrium (and an intermediate-run equilibrium), this is really the consequence of the mathematical relationship between the marginal and the average given the definition of elasticity.[12] Equation [5.5] shows that there is no formal difference between the returns to scale of the production function (its closeness to constant returns) and the elasticity of the firm's demand curve *in long-run equilibrium*.

Again we can see how *special* the linear-homogeneous production function is. Proposition [A] is consistent with [B] and [C] – that is, with a long-run equilibrium – but this is true only when $\varepsilon = -\infty$ (that is, when the price is given, $MR = AR \equiv$ price). Equation [5.5] shows this by noting that in this case $\beta = \infty$ or $(1/\beta) = 0$ which implies that the production function is (at least locally) linear-homogeneous.

Finally, note that the existence of 'increasing returns' is often called the case of 'excess capacity' – that is, where the firm is not exploiting the full capacity of its (fixed) plant which if it did it could lower its average cost (in other words, it is to the left of the lowest point on its *AC* curve). All this leads to the conclusion that when [D] holds with profit maximization, that is, with [B], either we have 'excess profits' (viz. when there are constant returns to scale) or we have 'excess capacity' (viz. when $TP = 0$).

PROFIT MAXIMIZATION [B]

Note that so far we have always assumed profit maximization. Let us now consider circumstances under which [B] does *not* hold. First let us assume that the firm is a perfect competitor, that is, that [D] does not hold. But this time we will assume the firm in the intermediate run is maximizing the 'rate of return' (r) on its capital[13] or what amounts to the same thing, is maximizing the average-net-product of capital (ANP_K) which is defined as,

$$ANP_K \equiv [X - (W/P_x)\cdot L] / K.$$

And since average productivity of capital (APP_K) is simply X/K,

$$ANP_K \equiv APP_K - (W/P_x)\cdot(L/K).$$

Moreover, when ANP_K is maximized in the intermediate run, the following holds:[14]

$$MPP_K = [X - (W/P_x)\cdot L] / K \equiv ANP_K \qquad [5.6]$$

$$MPP_L = W/P_x. \qquad [5.2a]$$

First let us see what this means if we assume [A] holds but not [C], such as when $TP > 0$. From the definition of TP, TR and TC, when $TP > 0$ we get:

$$P_x\cdot X > W\cdot L + P_k\cdot K$$

or, rearranging,

$$[X - (W/P_x)\cdot \mathrm{L}] / K > P_k/P_x.$$

Since by [5.6] the left side of this last inequality is equal to MPP_K if the firm is maximizing ANP_K, the firm cannot also be maximizing profit with respect to capital (because $TP \neq 0$). However, had we assumed that $TP = 0$, we would get the same situation as if [5.2a] and [5.2b] were the governing rules rather than [5.2a] and [5.6]. That is to say, if we assume the firm is in a long-run equilibrium, it does not matter whether the firm is a profit maximizer (i.e. [5.2b] holds) or thinks it is an ANP_K maximizer (i.e. [5.6] holds) with respect to capital. Now earlier we said that if [A] but not [D] holds the intermediate run implies a long-run equilibrium. Thus, if we only know that $TP > 0$, we can say that whenever [A] holds, [B] cannot hold except when [D] also holds. Alternatively, when $TP > 0$ whenever [D] does not hold, [A] cannot hold if [B] does.

ON BUILDING MORE 'REALISTIC' MODELS OF THE FIRM

Now all this leads us to an argument that we should avoid assuming linear-homogeneous production (i.e. assumption [A]) and thereby allow us to deal with the intermediate-run equilibrium with or without profit maximization. In particular, I think a realistic model of the firm will focus on the properties of an intermediate-run equilibrium which is not a long-run equilibrium, or on the excess capacity version of imperfect competition, both of which require that the firm's production function not be everywhere linear-homogeneous. Neither assumption denies the possibility that the production function can be *locally* linear-homogeneous at one or more points. This latter consideration means that the intermediate-run view of the firm offers the opportunity to explain *internally* the size of the firm in the long-run equilibrium. Size is impossible to explain if [A] holds (unless we introduce new ideas such as the financial endowments of each firm). Furthermore, it is again easy to see that competition is unimportant when [A] is assumed to hold and [D] does not. That is, the traditional argument that 'competition' is a good thing would be vacuous when [A] and [B] hold but [D] does not hold. This is because [A] and [B] alone (i.e. without the additional assumption that competition exists) imply [C] which was one of the 'good things' explicitly promised by long-time advocates of free-enterprise capitalism or more recently implicitly by advocates of the privatization of government-owned companies. So, again, if economists are to argue that competition matters, they must avoid [A].

USING MODELS OF DISEQUILIBRIUM

Now with the above elementary axiomatization of the Marshallian theory of the firm in mind, let us return to the consideration of how such a theory can be used to explain states of disequilibrium. To do this we need only consider each of the four models we will get when we decide which of the assumptions [A] to [D] we will relax (since, as I explained, the four assumptions cannot all be true simultaneously).

Model 1. Dropping assumption [D]

Dropping the notion that the firm can affect its price (by altering the quantity it supplies to the market) merely yields the old Marshallian theory of the price-taking firm (see Figure 5.1). Nevertheless, it does give us the opportunity to explain various states of disequilibrium. Let us consider various attributes of disequilibrium. If the firm is *not* at the point where the production function is locally linear-homogeneous, there can be several interpretations of the situation depending on whether or not we assume [B] or [C] holds. If [C] does not hold but [B] does, there could be either positive or negative profits. If we wish to explain the absence of zero profits, we can always claim that this is due to our not allowing sufficient time for competition to work. If [B] does not hold but [C] does, then there must be something inhibiting the firm from moving to the optimum point where price equals marginal costs. In comparative-statics terms, we can explain either type of disequilibrium state by noting that since the last state of equilibrium was reached certain exogenous givens have changed. For example, tastes may have changed in favour of one good against another, thus one firm will be making profits and another losses or the firm has not had enough time to move along its marginal cost curve. Similarly, it could be that technology has changed. Any such explanation thus would have to be specific about the time it takes to change variables such as capital as well as specify the changes in the appropriate exogenous variables. Hopefully, such an explanation would be testable.

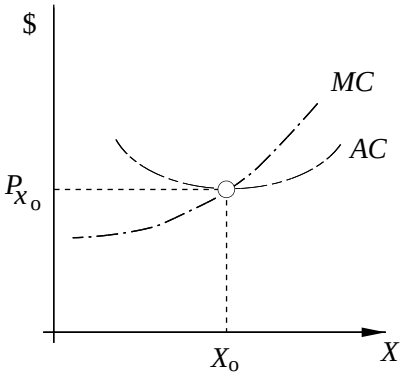

Figure 5.1 *Firm in long-run equilibrium*

Sometimes there is little difference between models which explain the occurrence of a disequilibrium phenomenon and those which explain it away. For example, models which drop assumption [D] usually explain away apparent disequilibrium phenomena as possible consequences re sulting from the limited amount of time available for competition to produce either zero profit or the optimum use of all inputs. The phenomena are suboptimal only in comparison with long-run equilibrium. Once one recognizes that there has not been enough time, as long as the firm is maximizing with respect to every *variable* input, nothing more can be expected. In other words, disequilibrium phenomena may be long-run disequilibria and short-run equilibria.

Model 2. Dropping assumption [B]

Dropping assumption [B] leads us astray from ordinary neoclassical models since [B] says that the firm is a maximizer. What we need to be able to explain is the situation depicted in Figures 5.2(a) and 5.2(b), again depending on whether or not we are assuming a long-run situation. In either case it is clear that the firm is setting price equal to marginal cost[15] which means that MPP_L equals W/P_x and thus cannot be satisfying equation [5.2b′] which is the necessary condition for profit maximization when [D] holds. An exception is possible if we assume the owner of the firm is not very smart and attempts to maximize the rate of return on capital rather than profit. For a maximum ANP_K, all that would be required is that ANP_K equals MPP_K. There is nothing inconsistent since it is still possible for [D] and [A] to hold so long as ANP_L equals MPP_L and this is the case. But again, maximizing rates of return to either labour or capital is not what we would normally assume in a neoclassical explanation.

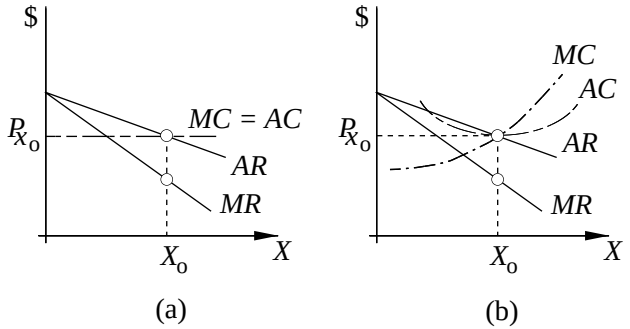

Figure 5.2 *[A] + [C] + [D] implies not-[B]*

Models which drop assumption [B] usually resort to a claim that there is some sort of unavoidable market failure or governmental interference preventing the firm from choosing the optimum amounts of inputs. Some imperfectly competitive firms are regulated to charge full-cost prices, that is, set price equal to average cost. Again, the apparent disequilibria may

still be the best that is possible. Since one cannot give a neoclassical explanation without assuming [B], one must resort to non-economic considerations such as external politics or internal social structure to explain the constraints that inhibit the firm from using the optimum amounts of inputs.

Model 3. Dropping assumption [A]

The most common disequilibrium model would involve the phenomenon of 'excess capacity'. The typical model is shown in Figure 5.3. There is no literal long-run version since if all inputs were variable (the definition of the long run) then [A] would have to hold. Models which drop assumption [A] usually try either to explain why excess capacity may be an optimal social equilibrium or to explain [D] away so that [A] can be allowed to hold. When [D] holds, competition can drive profits to zero without forcing the firm to a point where it faces local linear homogeneity. To see this we need only note that [B] combined with [C] is represented by equations [5.2a′], [5.2b′] and [5.3]. And as we noted before these imply that the firm is facing a falling *AC* curve since it must be facing increasing returns. As I noted above, the common justification of [D] is to say there are transaction costs which if recognized would explain that the situation represented by Figure 5.3 is an optimum rather than a disequilibrium. It is the best *possible* world.

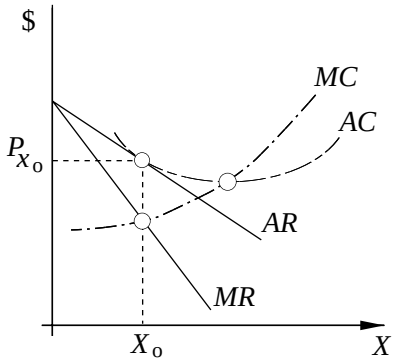

Figure 5.3 *Imperfectly competitive firm in long-run equilibrium*

Some people wish to interpret excess capacity as evidence that imperfect competition leads to inefficiencies where it is clear that the firm is not maximizing its output for the resources used (i.e. *AC* not minimum). It could equally be argued that the transaction costs needed to make decisions when there is the very large number of producers required to make everyone a perfect competitor are too high. A long-run equilibrium with zero profits and increasing returns may very well be the best we can do for society. Too often the transaction costs are invisible or imagined. The cleverest models are those which claim that the prices we see do not represent the true costs of purchase. The fact that people are willing to join a queue and wait to be served when there are few producers is interpreted as evidence that the price marked on the good is less than the price paid. The full price includes that opportunity cost of waiting (i.e. lost income). Thus, implicitly, the demand curve for the 'full' price is horizontal and the resulting 'full' cost curves if visible would look like Figure 5.1, thereby denying [D] and allowing [A] to be re-established. I think such a model may be too clever since it is difficult for me to understand what is being explained with such a model.

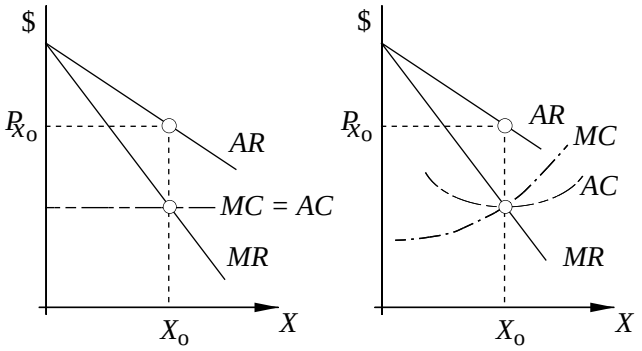

Figure 5.4 *[A] + [B] + [D] implies not-[C]*

Model 4. Dropping assumption [C]

One obvious way to explain the existence of profits is to simply drop [C] without dropping assumption [D]. The explanation in this case will be direct since given assumptions [A] and [B] it is logically impossible for profits to be zero or negative whenever [D] holds, hence the absence of zero profits is quite understandable. Consider Figures 5.4(a) and 5.4(b). In each figure we represent [D] by a falling demand curve (the *AR* curve) and its resulting marginal revenue curve which is necessarily always below. Assumption [B] is represented by the point where marginal revenue equals marginal cost. Assumption [A] is represented only at the point or points where average cost equals marginal cost. Which of Figures 5.4(a) or 5.4(b) is the appropriate representation depends on why [C] does not hold.

Models which initially drop assumption [C] will usually be transformed

into ones where [A] or [D] does not hold so that [C] can be allowed to hold. When the objective is to explain [D] away (e.g. with the recognition of 'full' costs), then [A] will be explained or explained away using one of the strategies I noted in the discussion of Model 3 and this leads to the re-establishment of Figure 5.1. Another strategy is to try to explain the *appearance* of profit as a return to an unrecognized input factor such that, when accounted for as a cost, total profit is really zero. This latter strategy allows [D] to hold but puts [A] or [B] into question. However, if there is only one missing factor, its recognition begs the question as to whether it is being optimally used. Only if [D] is denied can it be argued that the existence of profit implies that some of the factors are not being used optimally.

Simply assuming [C] does not hold may provide the logic necessary to explain profits, but if the firm operates in a competitive industry something needs to be added to explain why profits are not zero. Figure 5.4(a) would be appropriate if the reason given is that there has not been sufficient time for competition to force profits down to zero. If there has been enough time, then Figure 5.4(b) is appropriate since implicitly it is assumed that the firm is in the long run. If the firm is in the long run then there must exist exogenous barriers to inhibit entry or competition. One obvious way to justify that [C] does not hold is to deny the existence of sincere competition. Perhaps it is a matter of collusion. Perhaps it is a matter of high cost of entry. Perhaps it is a matter of government-imposed barriers to entry such as we sometimes see in the case of utilities (e.g. power utilities, telecommunications, transportation, broadcasting, etc.). Perhaps it is because of the exercise of power granted in the social setting of a firm, so-called exploitation of workers by the owners of the firm [see Robinson 1933/69].

Whatever the reason given, least-cost production [A] combined with maximization [B] means that the existence of a falling average revenue precludes negative profits. In other words, we can never explain a disequilibrium that involves negative profits with an imperfectly competitive neoclassical model based on [A] and [B]. Moreover, we are also limited to using such a model only to explain part of the economy since it is impossible to have an economy where everyone is making profits.[16] Aggregate profit for an entire (closed) economy must be zero, hence if any firm is making profits, some other firm must be making losses. Thus, the disequilibrium state of an entire economy cannot be explained with an imperfect-competition-based neoclassical model.

UNIFORMITIES IN EXPLANATIONS OF DISEQUILIBRIA

I will consider how many of the above models can be seen as variants which use the same mathematical property inherent in disequilibrium states. In one sense I have already discussed the notion that increasing returns and imperfect competition are two ways of interpreting what is represented in Figure 5.3. And I showed that in this case the measure of distance from the perfect competition equilibrium is either a measure of closeness to constant returns or a measure of closeness to perfectly elastic demand. The measures are equivalent.

Can we do something similar for all disequilibrium models? That is, are all explanations based on positing disequilibrium phenomena (inefficiency, exploitation, suboptimal resource allocations, profits, etc.) reducible to statements about some measure from the perfectly competitive optimum equilibrium?

Interest rate as a measure of disequilibrium

Let us examine some models which are based on the presumption of a state of disequilibrium. Many years ago, Oscar Lange [1935/36] presented an elaborate model which in effect claimed that the interest rate (actually, the net internal rate of return) is implicit in a firm's or economy's misallocation of resources between the production of final goods X (by firm x) and intermediate goods K (which are machines produced by firm m).[17]

Lange's Model

Let the economy consist of two firms which are given the following production function for final goods:

$$X = F(L_x, K_x) \qquad \text{[L1]}$$

and the following production function for machines which last only one production period:

$$(K_m + K_x) = \phi(K_m, L_m) \qquad \text{[L2]}$$

where the subscript indicates which firm is using the machine. And we note that [L2] also indicates that it will be assumed that the supply of machines is exactly equal to the demand for machines (which are assumed to be used up in one production period). Similarly, it will be assumed that the market for labour is cleared (i.e. there is full employment):

$$L = L_x + L_m. \qquad \text{[L3]}$$

Let us now assume the economy is producing with an allocation of labour between the two firms such that X is at its maximum. This assump

tion implies that there must be no surplus machine production on the margin (i.e. the last machine produced is used to replace the last machine used up):

$$(MPP_K)_m = 1 \qquad [L4]$$

and that there is an efficient resource allocation (i.e. $MRTS_x = MRTS_m$):

$$(MPP_L)_x / (MPP_K)_x = (MPP_L)_m / (MPP_K)_m. \qquad [L5']$$

Note that when [L4] holds with [L5′] it gives:[18]

$$(MPP_L)_x = (MPP_K)_x \cdot (MPP_L)_m. \qquad [L5]$$

If X is not maximum, either [L4] or [L5′] does not hold (or neither holds).

If we assume [L5′] holds because the two firms have somehow achieved an efficient allocation of labour between them, that is, they have achieved a Pareto optimum for the given amount of labour, L, then failure to maximize X must imply that equation [L4] does not hold. If the failure to maximize X is the result of misallocating too much labour to the production of X, then we can measure the extent to which [L4] does not hold by a scalar i as follows:

$$(MPP_K)_m = 1 + i. \qquad [L17]$$

This i is equivalent to what Lange calls a net 'rate of real interest'. Note that whenever this two-firm economy is *not* maximizing X but has reached a Pareto-optimal equilibrium in the sense that neither firm can increase its output without the other firm decreasing its output, i cannot be zero.[19] In other words, i is a measure of the distance the Pareto-optimal point is from the global optimum of a maximum X for the given amount of labour being allocated between these two firms.

We can look at Lange's real interest rate as a measure of increasing returns if we assume the machine producing firm is a profit maximizer. In effect equation [L17] can be the equivalent of my equation [5.2b′] once we recognize that the real price of capital in the production of machines is P_k/P_k thus [L17] is really:

$$(MPP_K)_m = (P_k/P_k)\cdot(1 + i). \qquad [L17']$$

Thus we can say that

$$(1 + i) = 1 / [1 + (1/\varepsilon)].$$

Since ε is in general a measure of the difference between the marginal and the average[20] (and thus equal to $-\beta$), we can determine the one-to-one correspondence between i and my measure of closeness to local linear homogeneity as follows:

$$(1 + i) = 1 / [1 - (1/\beta)]$$

or, equivalently, we can say either that

$$-i = 1 / (1 - \beta)$$

or that

$$-\beta = 1 - (1 / i).$$

Other measures of disequilibrium

Let us now consider other, more familiar or more recent, models of disequilibrium which claim to offer measures of the extent of disequilibrium and see whether we can generalize the relationship between those measures and either my β or equivalently the elasticity of demand. We will look at Robinson's [1933/69] measure of exploitation due to monopoly power, John Roemer's [1988] more general measure of exploitation, Abba Lerner's [1934] index of monopoly power, Michal Kalecki's [1938] degree of monopoly, and Sidney Weintraub's [1949] index of less-than-optimum output.

Robinson's measure of exploitation due to monopoly power is the difference between the marginal product of labour and the price paid for the labour services. This index can be derived straight from equation [5.2a′] above. In effect her measure is merely $1/\varepsilon$ since this fraction is the measure of the difference.

Roemer's measure of exploitation is the ratio of profit to variable costs. Roemer's measure does not assume [C] holds. If we assume that his disequilibrium model has only one input, then his measure is just

$$(\text{price} - AC)/AC.$$

If we also assume Roemer is presuming maximization in the sense that price equals MC then his measure of exploitation is just $1/\beta$.[21]

Kalecki's degree of monopoly is based on an assumption that [A] and [B] hold but [C] does not. Thus his measure is the difference between AR and MR which again is $1/\varepsilon$.

Lerner's index of monopoly power is defined as the ratio of difference between the price and MC as a proportion of the price, or since AR is price:

$$(AR - MC) / AR.$$

If we assume zero profit then his index is my $1/\beta$ and if instead we assume profit maximization ($MR = MC$), then his index is the negative of $1/\varepsilon$. If we assume both conditions hold (i.e. an imperfect competition equilibrium) then his index is equivalent to both my $1/\beta$ and $1/\varepsilon$ (as I explained earlier).

Weintraub's index of less-than-optimum output is the ratio of less-than-optimum output to optimum output where the optimum is the one where [A] holds or, equivalently, where $MC = AC$. Thus his index is dependent on the specific form of the production function or, equivalently, of the cost function. To illustrate, let us assume the total cost (TC) of producing X is as follows:

$$TC = 200 + 10X + 2X^2$$

then $$AC = (200 + 10X + 2X^2) / X$$

$$MC = 10 + 4X.$$

Now let us calculate the ratio of MC to AC using the given cost function:

$$MC / AC = X\cdot(10 + 4X) / (200 + 10X + 2X^2)$$

or $$MC / AC = (5X + 2X^2) / (100 + 5X + X^2).$$

Note that $MC = AC$ when $X = 10$ and thus Weintraub's index (WI) will be (X/10) for the given cost function. Since $MC = AC\cdot[1 - (1/\beta)]$, we can calculate β *for the given cost function* if we are given an X:

$$\beta = (6 + WI + 2WI^2) / (2WI^2 - 6).$$

So, again, we see that the measure of distance from a perfectly competitive equilibrium can be seen as a variant of β or ε.

A GENERAL THEORY OF DISEQUILIBRIA

In general terms, each of the models of disequilibrium I have discussed here are combinations of the axioms I have presented in this chapter. Which of the four axioms ([A] to [D]) is denied will be the basis for a clearly defined measure of disequilibriumness. The opportunities for criticism are limited to examining the reasons why the particular axiom was denied. And since any measure of disequilibrium will be determined by the denied axiom, not much will be learned by arguing over the nature of the measure presented. In general, unless the same axioms are used to build alternative models of disequilibrium, arguing over which is a better measure would seem to be fruitless. Whether the disequilibriumness is the result of assuming [D] or [A] in combination with either [B] or [C] will determine which is the appropriate index. And as we saw in the case of imperfectly competitive equilibria, either index will do. With the one exception of Kalecki's degree of monopoly which neutralized the role of the production function by assuming linear homogeneity [A], all of the other measures can be seen to depend on the extent to which the production function is not linear-homogeneous (as measured by my β).

The questions of the pervasiveness of equilibrium and maximization are fundamental and thus little of neoclassical literature seems willing or able to critically examine these fundamental ideas. Outside of neoclassical literature, however, one can find many critiques that are focused on what are claimed to be essential but neglected elements of neoclassical explanations. There are two particular exogenous elements that have received extensive critical examination. One is the question of what a decision-maker needs to know to be a subject of the maximization assumption. The other involves the social institutions that are needed yet taken for granted in neoclassical explanations. The critics complain that until these two exogenous elements are made endogenous, neoclassical theories will always be incomplete. While some critics argue that such a completion is impossible, some friends of neoclassical theory willingly accept the challenge. In the next three chapters I will examine these disputes to determine the extent to which they represent serious challenges to neoclassical economics.

NOTES

1 There have been some analyses of the stability of equilibrium models which recognize the need to deal with conceivable disequilibrium states [e.g. Hahn 1970; Fisher 1981, 1983]. Also, in macroeconomics we find models which try to deal with the disequilibria caused by 'distortions' such as sticky prices or wage rates [e.g. Clower 1965; Barro and Grossman 1971]. Little of this literature approaches the way equilibrium models have been axiomatized. Besides, it is not clear what consistency and completeness mean when one sees disequilibrium as a distorted equilibrium.

2 It might appear that by assuming all consumers are maximizing we are always assuming that the only possible disequilibrium is one of excess supply, that is, for disequilibrium prices above the equilibrium level. This does not have to be the case if one adopts the Marshallian view of the producer where the given price is a demand price and marginal cost represents the supply price. In this way, prices on both sides of the equilibrium level can be considered.

3 Here 'capital' always refers to physically real capital (e.g. machines and computers, etc.).

4 If all inputs are unrestricted then it is possible to double output either through internal expansion (viz. by doubling all inputs) or through external expansion (viz. by building a duplicate plant next door). It should not matter which way. If it does matter then it follows that not all inputs are variable. By definition, a linear-homogeneous function is one where it does not matter which way output is expanded. Some of my colleagues argue that, even in the long run, some production functions cannot be linear-homogeneous. They give as an example the production of iron pipe. One can double the capacity of the pipe without doubling the amount of iron used – the perimeter of the pipe does not double when we double the area of the pipe's cross-section. Unfortunately, this example does not represent a counter-example as claimed. To test linear homogeneity one would have to restrict consideration to producing more of the *same* product and 20-inch pipe is not the same product as 10-inch pipe.

5 It should be noted that equations [5.1], [5.2a], [5.2b] and [5.3] are formalizations of the statements (b) to (d) used to discuss Marshall's method (see above, pp. 32–5).

6 That is, if [5.1], [5.2a] and [5.3] hold, [5.2b] must also hold.

7 That is, $MC \equiv W/MPP_L$.

8 The calculation follows from the definitions of these terms:

$$\varepsilon \equiv (\partial Q/Q)/(\partial P/P) \equiv (P/Q)\cdot(\partial Q/\partial P)$$

and

$$MR \equiv \partial(P\cdot Q)/\partial Q \equiv Q\cdot(\partial P/\partial Q) + P\cdot(\partial Q/\partial Q)$$

$$\equiv P\cdot[1 + (Q/P)\cdot(\P P/\P Q)].$$

Thus,

$$MR \equiv P\cdot[1 + (1/e)]$$

and since $P = AR$, the relationship between AR and MR follows.

9 See above, p. 66.

10 The implausibility of the firm being a perfect competitor with regard to *output* prices does not necessarily imply an implausibility of the firm being a perfect competitor with respect to *input* prices. That is, a few big firms in one industry still may compete with many other industries for labour (or capital). This, of course, assumes at least a minimum degree of homogeneity or mobility of labour – that labour could easily move from one industry to another. If for any reason this is not the case, then we will have to include the elasticity of labour supply, x, in the calculation of Marginal Cost. If we do this, we will get (for the short-run equilibrium):

$$MC = (W/MPP_L)\ \ [1 + (1/x)]$$

But since I wish to keep things as uncomplicated as possible here I will not develop this type of imperfect competition further.

11 The difficulties with combining the notion of imperfect competition with a long-run or general equilibrium model are not new. Recent discussion [e.g. Hart 1985; Bonanno 1990] have complained that most attempts to do so [e.g. Negishi 1961] usually have involved compromising assumptions that leave the end results far from being an ordinary general equilibrium model augmented with the assumption of imperfect competition. John Roberts and Hugo Sonnenshein [1977] seem to be going further by arguing that such an augmentation is impossible. In my simple-minded arguments which follow it seems that the problem is not just a question of coming up with a clever modelling technique but rather a fundamental logical obstacle.

12 That is, by analogy we can see that using equation [5.5] yields:

$$-b = (\P Q/Q)/(\P AC/AC) \equiv (AC/Q)\cdot(\P Q/\P AC)$$

and since $MC \equiv \P(AC\cdot Q)/\P Q \equiv Q\cdot(\P AC/\P Q) + AC\cdot(\P Q/\P Q)$

$$\equiv AC\cdot[1 + (Q/AC)\cdot(\P AC/\P Q)]$$

we get $$MC \equiv AC\cdot[1 - (1/b)].$$

13 Consideration of the *intermediate-run* equilibrium makes it possible to entertain an alternative assumption for the goal of the firm in the intermediate run even when the firm may wish to maximize profit in the short run. While it will be easy to show that maximizing the rate of return makes sense only when comparing equal amounts of investment (i.e. it is possible to make more profit at a lower rate of return when the amount is not fixed), it is not uncommon to find people bragging about high rates of return achieved as if this were optimal.

14 Consider the relationship between MPP_K and ANP_K. In particular, let us show that

$ANP_K = MPP_K$ whenever ANP_K is maximum (with respect to K), and

$ANP_K < MPP_K$ whenever ANP_K is rising as K increases.

By definition:

$$ANP_K \equiv [X - (W/P_x)\cdot L] / K. \qquad [i]$$

Now let us determine the slope of the ANP_K curve $(\P ANP_K/\P K)$ by differentiating equation [i]:

$$(\P ANP_K/\P K) = [(\P X/\P K) - 0]/K + [X - (W/P_x)\cdot L]\cdot(-1)\cdot(\P K/\P K)/K^2. \qquad [ii]$$

Since by [i]:

$$ANP_K\cdot K \equiv X - (W/P_x)\cdot L$$

we can transform [ii] into the following:

$$(\P ANP_K/\P K) = [(\P X/\P K)/K] - [(ANP_K\cdot K)\cdot(\P K/\P K)]/K^2. \qquad [iii]$$

Since $(\P X/\P K) \equiv MPP_K$ and $(\P K/\P K) = 1$, we can further obtain:

$$(\P ANP_K/\P K) = (MPP_K - ANP_K)/K. \qquad [iv]$$

With [iv] we can see that if the slope is positive (i.e. ANP_K rising) then $(MPP_K - ANP_K) > 0$, which implies $MPP_K > ANP_K$. And, if the slope is zero (i.e. the slope is horizontal when ANP_K is maximum) then $(MPP_K - ANP_K) = 0$, which implies $MPP_K = ANP_K$.

QED

15 Note that the marginal and average cost curves are short-run curves in Figure 5.2(b). I will not try to define an intermediate version since it will not add much to the analysis.

16 As Samuelson [1972] noted, for there to be a net profit for an entire economy begs the question of whether there is a Santa Claus [see further Boland 1986a, Chapter 2].

17 Lange uses m to represent the output of machines but here I will use K, to maintain the notation of this chapter.

18 It should be noted here that Lange does not state equation [L5¢] since he derives both [L4] and [L5] using Lagrange multipliers and thus implicitly assumes [L4] and [L5] are both true. By recognizing [L5¢], I am making it possible to treat [L4] and [L5¢] separately while still recognizing that Lange's equation [L5] is also a necessary condition of a maximum X.

19 According to Lange, the real rate of interest is zero when X is maximum [p. 169]. It should be noted here that my representation of Lange's model is slightly different from what he explicitly states. Lange takes equation [L4] as obviously true such that any disequilibrium can *only* be the result of my equation [L5¢] not being true. All of Lange's propositions still follow from my representation of his model.

20 See note 8.

21 If instead we assume the profit-maximizing firm has two inputs, L and K, then the measure $(1/b)$ is increased by the factor $[1 + (P_k\cdot K)/(W\cdot L)]$.

Part II

Some neglected elements