

Money, Employment and Inflation

26.1 THE DEVELOPMENT OF THE KEYNESIAN SYSTEM

The reception of the General Theory

Keynes' *General Theory*¹ was the outcome of two decades in which an enormous amount of work was put into macroeconomic theorizing. For all Keynes' claims to novelty, much of the *General Theory*'s contents can be found elsewhere in the literature of the 1920s and 1930s, in particular in the writings of Pigou, Hawtrey and the Swedish school. Despite this continuity, however, discussions of modern macroeconomics have to start with the *General Theory*, the reason being that certain interpretations of it came to dominate post-war macroeconomics. Earlier contributions, irrespective of their merits, were almost completely eclipsed.

The *General Theory* was immediately successful, the extent and speed of its impact being unparalleled: the following year's *Economic Journal* was, perhaps not surprisingly, filled with reactions to the book; a symposium on it filled an issue of the 1937 *Quarterly Journal of Economics*; and within a couple of years the framework provided by the *General Theory* was being used not simply by Keynes' colleagues, but by complete outsiders and even by economists initially critical of it. One historian has made the claim that "During the ten or twelve years after its appearance the *General Theory* received more attention than Alfred Marshall's *Principles of Economics* had received in over 50 years."² A survey of contemporary economics published at the end of the 1940s indicates the extent of his influence: he was cited 98 times, compared with 56 citations for Hicks, his nearest rival. In addition, ideas stemming from the *General Theory* formed the subject matter of at least part of 15 out of the 23 chapters in the survey.³

To a certain extent the reception of Keynesian economics can be seen in terms of the emergence of a new generation of economists, the disciples of Keynes (e.g. Samuelson, Robinson, Harrod, Meade, Lerner) belonging to the younger generation, the critics (e.g. Leontief, Schumpeter, Knight, Viner) to the older generation.⁴ This still leaves open, however, the question of why Keynesian ideas were found so attractive. One reason was undoubtedly that Keynes was addressing an important contemporary problem: though recovery from the depression was by 1937 well under way, the depression of 1929 to 1933 had had a profound impact. Samuelson has argued that although it still survived, events since 1929 had caused belief in the existing orthodoxy to atrophy, and that the *General Theory* provided an alternative.⁵ Also important was the theoretical challenge provided by

the *General Theory*, for not only did it contain a wealth of ideas capable of development, but it also contained many puzzles requiring solution. Some economists have gone so far as to claim that its complexity was an essential ingredient of its success.⁶

With hindsight, however, it is perhaps the arguments used in criticism of the *General Theory* which are most interesting, for it is easy to see its attractions. Consider the arguments used by Schumpeter and Leontief, both of whom criticized Keynes on methodological grounds. Leontief concentrated on Keynes' method, shared with other members of the "Cambridge school", which Leontief described as "implicit theorizing".⁷ By implicit theorizing Leontief meant the practice of defining terms in such a way as to imply a definite theoretical relationship between these terms and the basic postulates of the theory, but without ever specifying this relationship.⁸ Such a procedure, according to Leontief, made rational discussion of a theory very difficult, for it was never clear to a critic exactly what the theory involved. An example was Keynes' use of aggregate demand and supply curves without specifying the factors on which they depended sufficiently carefully for it to be possible to say whether or not it was correct to treat them as independent of each other. He agreed with Schumpeter's accusation that Keynes shared what Schumpeter described as the Ricardian vice: "the habit of piling a heavy load of practical conclusions upon a tenuous framework, which was unequal to it yet seemed in its simplicity not only attractive but also convincing". Schumpeter saw this as having dramatic implications, for not only did Keynes influence the best minds of the economics profession, but he also "brought back the happy times of Mrs. Marcet, when every schoolgirl, by learning the use of a few simple concepts, acquired competence to judge all the ins and outs of the infinitely complex organism of capitalist society".⁹ The simplifications to which Schumpeter objected included the theory's static and short run nature; its neglect of the effects of investment on capacity; and the assumption that people respond to real values everywhere except when it comes to bargaining over money wages. It was only with all the givens implied by these assumptions that "the three great simplifiers", as Schumpeter called the consumption function, the marginal efficiency of capital and liquidity preference, could determine national income and implement Keynes' vision of the economic system.¹⁰

It was in the course of the debates over the *General Theory* that the contributions of the Swedish economists were brought to a wide audience. This was done by Ohlin (1937), who coined the phrase "the Stockholm school" to describe the work of Lindahl, Myrdal and himself; work which he saw as anticipating the essential ideas of the *General Theory*.¹¹

Extensions of the General Theory

Much of the literature appearing in the decade after 1936 was concerned with clarifying and debating detailed points made in the *General Theory*: liquidity preference versus loanable funds theories of the interest rate; the

value of aggregative economics; the nature of the multiplier; the role of hoarding in the theory of liquidity preference. These controversies were, in the words of one historian, "long and tedious",¹² for though there were many points at which Keynes' arguments needed correcting, and though some important modifications were introduced (such as the *ex ante/ex post* distinction), they did not result in any fundamental changes to the Keynesian theory.

Of more interest are the investigations into the individual functions used in the *General Theory*. The notable success here concerned the consumption function, where Keynesian theory was examined in the light of empirical data, especially in the US. Studies of family budgets proved consistent with Keynes' consumption function, but against this evidence were cyclical variations in the propensity to consume, and evidence that in the long run the consumption function was, for the US at least, shifting upwards. This proved an ideal area for development, the most well-known theories being those of Duesenberry (1948), Friedman (1957) and Modigliani and Brumberg (1954). Of these the one that took firmest root was the permanent income/life cycle theory (it can be argued that the two are substantially the same). The reason usually given for the success of this theory¹³ is that the Keynesian hypothesis was refuted by the evidence, the permanent income/life cycle theory replacing it because of its ability to explain this evidence. This is, however, only part of the story, for it could be argued that the success of the permanent income/life cycle theory arose because it explained the consumption function in terms of individual maximizing behaviour: until it had been explained in these terms there was a sense in which it was not understood.

This interpretation of the theory of the consumption function is supported by developments in other areas, such as money and wages. Shortly after the *General Theory* evidence was produced to show that wages did not vary over the cycle in the way predicted by the *General Theory*.¹⁴ Though this added to the controversy surrounding the *General Theory* this did not stimulate new theoretical developments until the 1970s. Two reasons suggest themselves. The first is that the Keynesian system emphasized demand rather than supply, so the behaviour of money and real wages could be regarded as an anomaly, peripheral to the main theory. This changed in the 1970s. The other reason is that until "disequilibrium" models appeared, economists did not have a theoretical framework in which to analyse the problem: there was no microeconomic theory available to explain why firms should not employ labour to the point where the real wage equalled the marginal product of labour. The availability of suitable techniques was also important in attempts to improve on the Keynesian theory of the demand for money, where the main achievements were the theories of Baumol (1952) and Tobin (1958). The contribution of both these theories was to make sense of the demand for money, through providing an explanation in terms of individual maximizing behaviour. Thus it can be argued that, in these areas, a crucial factor was the desire to make sense of Keynesian ideas in terms of individual maximizing behaviour.

The *General Theory* was also extended by being applied to new areas: the business cycle (e.g. Harrod, 1936), growth (e.g. Harrod, 1939), the balance of payments (e.g. Robinson, 1937), and inflation (e.g. Keynes, 1940). These areas are discussed elsewhere in this chapter.

Keynes and the classics

In discussing the debates over the *General Theory* we have so far avoided the central issue – that of the nature of the *General Theory* itself. The tone of the debate was set by Keynes himself when he presented his theory as an alternative to a “classical” theory, his own theory being the more general. Controversy was aroused, not only because of Keynes’ claim to have provided a more general theory, but also because Keynes’ version of the classical theory was based on Pigou’s *Theory of Employment* (1933), a modern work which to many economists was as hard to understand as the *General Theory*. There were thus two aspects to the controversy: working out what the classical theory really was,¹⁵ and working out how it related to Keynes’ theory, for it was by solving these problems that economists hoped to sort out what were the essential features of the *General Theory*. Here the outstanding contribution was that of Hicks (1937, 1939a) whose interpretation of the *General Theory* became accepted to such an extent that it became, for many economists, not only synonymous with Keynesian economics, but the only framework within which macroeconomics could be conducted.

As mentioned above,¹⁶ Hicks came to the *General Theory* familiar with both Walrasian general equilibrium theory, and the writings of the Swedish economists. Furthermore he has already produced both a diagram with which to analyse three-way exchange, and an aggregative model involving labour, loans and output.¹⁷ Though in no way an anticipation of the *General Theory*, this led naturally into an interpretation of the *General Theory* as a temporary general equilibrium model, analysed in terms of IS and LM curves. The Swedish influence was most prominent in *Value and Capital* where expectations were treated more thoroughly than in his 1937 article, and where Myrdal’s *ex post/ex ante* distinction was used to sort out the relationship between saving and investment, Keynes’ exposition of this having caused great confusion.¹⁸ His 1937 article showed what could be done with the diagram, in particular using the liquidity trap and the horizontal LM curve to portray the *General Theory* as being concerned with the “economics of depression”, applying when the prospective attractiveness of investment is so low that the IS curve cuts the flat portion of the LM curve. A higher incentive to invest would cause the IS curve to cut the steep portion of the LM curve, producing “classical” results.¹⁹

The IS–LM apparatus permitted Hicks to relax certain of Keynes’ assumptions, inserting income in the investment function and the interest rate in the savings function. This was significant in that it enabled him to find a similarity between Keynes and Wicksell, relating Keynesian ideas to Wicksell’s distinction between the natural and money rates of interest.²⁰

Differences between Keynes and the classics were argued to depend on the slopes of the IS and LM curves, and hence on the elasticities of savings, investment and demand for money with respect to income and the interest rate.

This approach, of using a simple four-market general equilibrium model, dominated discussion of Keynesian economics until the 1960s, controversy settling on the specific assumptions necessary to produce Keynesian results: was unemployment caused by wage stickiness (in which case some of Keynes' critics would consider themselves vindicated) or could it be caused by a liquidity trap or insufficient and interest-inelastic investment demand? It was in this discussion that the "real balance effect" came to be seen as crucial, for it was this which undermined the claim that insufficient investment demand or a liquidity trap was sufficient to create the possibility of an unemployment equilibrium, independently of wage rigidity. If wages were not rigid, unemployment would cause deflation which, through raising the real value of people's money balances, would raise consumption and increase employment.²¹

The climax of this debate came with Patinkin's *Money, Interest and Prices*, published in 1956 with an extensively revised edition in 1965. This book did two things. It provided a microeconomic theory on which monetary theory could be based, "redoing" Hicks' *Value and Capital* with money treated properly as a "very special good".²² In addition it provided a definitive account, within the Hicksian framework, of the difference between Keynes and the classics. His conclusion was that a full employment equilibrium would exist and that it would be stable. Keynesian economics was the economics of disequilibrium, arising when the equilibrating forces were too weak to restore full employment "within a socially acceptable period of time".²³ Thus Keynes was seen not as having made a fundamental theoretical contribution so much as having drawn attention to the case relevant to policy-making, namely disequilibrium.

The economics of disequilibrium

The interpretation of Keynesian economics described above became the established orthodoxy, enshrined in macroeconomics textbooks.²⁴ This changed, however, in the late 1960s when it was challenged first by Clower (1965), and later by Leijonhufvud in a widely read book *On Keynesian Economics and the Economics of Keynes* (1968). As its title implies, the book's claim was that what was generally taken to be "Keynesian economics" (namely the Hicksian interpretation) was in fact nothing like the economics of Keynes himself. Clower's point, which was taken up by Leijonhufvud, was present in *Money, Interest and Prices*, but was not fully developed there. It was that the Walrasian model, used by Hicks and Patinkin, did not deal with supply and demand in a manner appropriate to a discussion of Keynesian economics. The basis for this claim was that when markets fail to clear, transactors face not only budget and technological constraints, but also constraints on how much they can buy and sell. These constraints will

affect demands and supplies in other markets. For example, if a household finds that it cannot sell as much labour as it wishes, it will have to demand a smaller quantity of consumption goods than it would otherwise demand. This explains why “Walras’s Law”, which states that the sum of realized excess demands must be zero, may fail: it is possible to have, for example, an excess supply of labour without there being any corresponding excess demand. It was on the basis of this argument that Clower and Leijonhufvud were able to claim that Keynes had made a fundamental contribution to the theory of value: he had analysed the behaviour of an economy in which markets do not necessarily clear.

Leijonhufvud buttressed this argument with many detailed arguments concerning aggregation, wage rigidity and expectations in order to show that the orthodox interpretation of Keynes was wrong. His central theme was that Keynesian economics was about what happened when there was no Walrasian auctioneer, and when quantities adjusted faster than prices in response to an imbalance between supply and demand. It is the absence of the auctioneer which means that transactions take place at disequilibrium prices, resulting in quantity constraints being imposed on buyers and sellers. Keynesian economics is thus about coordination failures which are inherent in the market mechanism, not about what happens when certain prices are held constant. It is thus Keynesian economics which is the more general case.

Clower and Leijonhufvud challenged the orthodox interpretation of Keynes, but neither of them provided an alternative, usable framework for doing short run macroeconomics. Indeed Leijonhufvud has, in the 1970s, reacted strongly against formal modelling of Keynesian economics, seeing the crucial aspects of the *General Theory* as lying at the “presuppositional level”, it being possible to appraise Keynesian economics only in terms of “an informal and improvised meta-language”, not in terms of a formal model.²⁵ A more formal analysis of Keynesian problems, embodying Clower’s insight, was produced by Barro and Grossman (1971), who provided what they called a “general disequilibrium model”.²⁶ Taking a model simplified so that only two markets (for labour and goods) need be considered, they showed the possibility of an equilibrium where output is low because demand for goods is low; and demand for goods is low because output, and hence income, is low. The significance of such an equilibrium is that it can arise even if the real wage rate is at its equilibrium level. In such a framework multiplier effects arise naturally. Though these “disequilibrium” models started off as attempts to provide theoretical explanations of Keynesian problems they did not stop there, for it proved possible to analyse other cases, in particular classical unemployment (caused by an excessive real wage rate) and repressed inflation (where there is general excess demand, both for goods and for labour). In addition such models could be used to explain the behaviour of wages over the cycle, to analyse the balance of payments (Dixit, 1978) and to analyse capital accumulation (Malinvaud, 1980). The basic ideas were also used at a microeconomic level.²⁷

These models were still, however, not completely satisfactory, for they all took prices as given, failing to answer the question of why prices did not adjust to ensure full employment equilibrium. This was something to which answers could not be found at a macro level.

26.2 MONEY AND INFLATION

The theory of inflation before 1958

In the wake of the *General Theory* inflation began to be explained in terms of the flow of spending relative to the flow of output, rather than in terms of the quantity (stock) of money. The application of Keynesian ideas to inflation was made by Keynes himself in *How to Pay for the War* (1940): inflation was there viewed as produced by the combination of high purchasing power (and hence high expenditure) caused by war production and the reduced availability of goods caused by the diversion of goods into the war effort. The crucial assumption was that, because the economy was operating at full employment, any rise in demand would be met by a rise in prices rather than by a rise in output. In advancing this theory Keynes revived the “widow’s cruse” theory of the *Treatise on Money*,²⁸ for he assumed that rising prices would cause a rise in profits, the shift in the distribution of income towards profits being the means through which output and expenditure were brought into balance. Such theories of inflation subsequently became known as “inflation gap” theories, inflation being caused by the “gap” between expenditure and the quantity of goods available to meet that expenditure, the most widely read account of such a model being that enshrined in Samuelson’s (1948a) textbook. These inflation gap models differed from Keynes’ in that there was less emphasis on changes in the distribution of income as the means whereby savings and investment would be brought into equilibrium: rising prices might reduce demand through money illusion, rising marginal tax rates, the real balance effect as well as or instead of through a change in income distribution.

A particularly interesting version of the inflation gap theory was Hansen’s (1951) “two gap” model. The interesting feature of this model was that it dealt with both goods and labour markets: excess demand for goods would cause changes in the price of output; whilst excess demand for labour would cause changes in the money wage rate. Equilibrium required that prices and wages rise at the same rate, and hence that the goods gap and the labour, or factor, gap be equal. If the goods gap were larger than the factor gap, for example, the real wage would fall and so output and employment would be increased, thus reducing the goods gap and increasing the factor gap, bringing them into equality. The significance of the model is that it gives a particular equilibrium inflation rate; and that it represents inflation as the outcome of processes affecting the economy as a whole: inflation results not from what happens in a single market, but from the interaction of all the markets in the economy.

The 1950s was the era when it was common to distinguish between “demand”, or “demand pull” and “cost”, or “cost push” inflation.²⁹ Demand inflation covered the inflation gap theories just discussed, and also theories such as that of “demand shift” inflation, according to which inflation arose because prices were inflexible downwards: if demand shifts from good A to good B the price of A will rise, whereas that of B will not fall, so the overall price level rises.³⁰ Cost inflation, on the other hand, resulted from market power being used to push up wages or profit margins, with monetary and fiscal policy accommodating to the resulting rises in prices. Much attention was concentrated on unions and wage bargaining as a source of inflation. Issues raised included the degree of competition, oligopoly being seen as conducive to inflation; union rivalry; attempts to maintain or create wage differentials; and the pricing policies of firms.

The Phillips curve

Of more importance for subsequent developments in the theory of inflation was the emergence of the Phillips curve as the standard tool with which to analyse the problem of inflation.³¹ The curve derived by Phillips (1958) was an *empirically determined* relationship between the rate of change of money wage rates and the rate of unemployment; Phillips claimed that a curve which he had derived from data for the period 1861 to 1913 would also explain data for both the inter-war and post-war periods. What gave this model its significance was that it could be explained theoretically in terms of a standard model of the labour market. Lipsey (1960a) provided such an explanation, taking the standard theory, according to which wage changes were proportional to excess demand for labour, and modifying it to allow for frictional unemployment and vacancies. This explained the position and the shape of Phillips’ curve. The Phillips curve was discussed in relation to the US by Samuelson and Solow (1960), who also examined its implications for anti-inflation policy. In the late 1950s and the early 1960s an enormous number of empirical studies appeared which confirmed the negative relationship between inflation and unemployment that Phillips had discovered.³²

The Phillips curve, in a sense, marked the demise of the distinction between demand and cost inflation. Though it could be used alongside the demand–cost distinction³³ this was unnecessary. Arguments could instead be conducted by incorporating more variables into the Phillips curve. One such attempt was that of Hines (1964), who inserted a measure of unionization to measure union power. This provoked much research, in the late 1960s and early 1970s, on appropriate measures of union power which could be inserted in a Phillips curve, either to prove or to disprove Hines’ contention that union power had a significant impact on the inflation rate.³⁴ Far more significant, however, was the introduction of inflationary expectations into the Phillips curve. The main contributions here were two papers appearing in 1967–1968. Friedman’s presidential address to the American Economic Association (1968) and Phelps (1967). The introduction of

inflationary expectations was not new (for example, Samuelson and Solow had in 1960 raised the possibility that if the economy were run at high levels of unemployment inflationary expectations might fall, shifting the Phillips curve down)³⁵ but its implications had not been fully worked out before. Phelps and Friedman both argued that in the short run a downward-sloping Phillips curve would emerge, but that in the long run any trade-off between unemployment and inflation would disappear.

Friedman's argument was that the Phillips curve was, from the point of view of economic theory, mis-specified in that because supply and demand for labour both depended on real wages it was the rate of growth of real wages, not money wages, which should depend on unemployment. It was for this reason that expected inflation should appear on the right hand side of the Phillips curve: because everyone was concerned with real wages, a 1% rise in the expected inflation rate should lead to a 1% rise in the actual rate. Crucial to Friedman's argument was the concept of the "natural rate of unemployment", the unemployment rate consistent with a constant rate of inflation. He described this as the unemployment rate

that would be ground out by the Walrasian system of general equilibrium equations, provided that there is embedded in them the actual structural characteristics of the labour and commodity markets, including market imperfections, stochastic variability in demands and supplies, the cost of gathering information about job vacancies and labor availabilities, the costs of mobility, and so on.³⁶

Unemployment would differ from this natural rate only if people made mistakes in their inflationary expectations: low unemployment would be the result of inflation being underestimated, high unemployment the result of its being overestimated.

The emphasis of Phelps' argument was rather different, but its conclusions were remarkably similar. Phelps was concerned, in a way that Friedman was not, with the microeconomics of the labour market: with the question of how labour markets operate when information is incomplete and costly. His argument was that information about wages would travel only slowly, the result of this being that in the short term a rise in demand would induce workers to supply more labour – when they received higher wage offers they would at first interpret this as meaning that they had been offered an unusually good wage rate; only later would they discover that wages generally had risen and that there was nothing special about the wage they had been offered. This approach was explored by Phelps and others in *Microeconomic Foundations of Employment and Inflation Theory* (1970), where all but one of the contributors used a version of the natural rate hypothesis.

These theoretical developments were reflected in empirical work, where, making the assumption that expectations could be modelled as a lagged function of past inflation rates (so-called "adaptive expectations") emphasis was on estimating the size of the coefficient attached to expected inflation: did a 1% increase in inflationary expectations cause inflation to rise by a full 1%, or would inflation rise by only a fraction of 1%?³⁷ This was important because if it was the latter then there would still be a trade-off between inflation and unemployment in the long run, albeit with a much steeper

Phillips curve than in the short run. If, on the other hand, the coefficient were unity then the natural rate hypothesis was vindicated. Initially empirical work suggested a coefficient significantly less than unity, with the result that the natural rate hypothesis was not accepted.³⁸ This attitude changed during the 1970s, however, when it became generally accepted that the long run Phillips curve must be vertical. One reason for this was empirical work: wage equations, perhaps due to the acceleration of inflation in the early 1970s, began to produce higher values for the coefficient on inflationary expectations. In addition, the relationship between unemployment and vacancies involved in the traditional approach appeared to break down. But more important were theoretical developments, in particular the arguments surrounding rational expectations. As long as the Phillips curve was thought of as something “tacked on” to a basically Keynesian model, there was no incongruity in having a non-vertical long run Phillips curve. As the Phillips curve became a more integral part of the model, a model supposedly grounded in the assumption of rational behaviour, it became natural to assume that the long run Phillips curve must be vertical. A symptom of this change was the replacement, in macroeconomics textbooks, of the IS–LM model with the aggregate supply and demand model as the main explanation of output and the price level.³⁹

Friedman and the quantity theory of money

In the wake of Keynesian economics the quantity theory of money was neglected as an explanation of inflation and output in preference to the income expenditure approach. The quantity theory was far from neglected in economic theory, however, where the literature contained extensive discussion of two issues: the “classical dichotomy”,⁴⁰ attacked by Keynes, according to which the real factors underlying supply and demand determined relative prices, with the quantity of money determining only the absolute price level; and the neutrality of money (whether an increase in the money supply will do anything other than raise all prices in the same proportion). Discussion was, however, at a purely theoretical level.⁴¹ The main contribution towards reviving interest in the quantity theory as a useful macroeconomic tool was Friedman’s article, “The quantity theory of money: a restatement” (1956). The central argument of this article is contained in the following quotation.

The quantity theory is in the first instance a theory of the *demand* for money. It is not a theory of output, or of money income, or of the price level. Any statement about these variables requires combining the quantity theory of money with some specifications about the conditions of supply of money and perhaps about other variables as well.⁴²

In formulating a theory of the demand for money Friedman stressed that it was one asset, one way of holding wealth, and that demand for it could be analysed using the standard theory of consumer choice. This led him to write the demand for money function as

$$M = f(r_b, r_e, (1/P) (dP/dt), w, Y/P, u)$$

where demand for money depends on the expected yields on holding bonds and equities, the expected inflation rate, the ratio of human to non-human wealth, real income (Y/P) and variables affecting tastes and preferences (summarized by u). Though the stress on price expectations and the inclusion of the ratio of human to non-human wealth (which caused demand for money to fluctuate over the cycle) differentiated Friedman's theory from that of Keynes, there is considerable justification for describing Friedman's theory as a more elegant statement of Keynes' theory of liquidity preference. The velocity of circulation is simply $1/f$.

The quantity theory, according to Friedman, comprised two assertions:

- (i) the empirical hypothesis that the demand for money is stable – more stable than functions such as the consumption function that are offered as alternative key relations.
- (ii) there are important factors affecting the supply of money that do not affect the demand for money.⁴³

Friedman presents his ideas as a restatement of an oral tradition existing at Chicago, where a version of the quantity theory was kept alive by Simons, Mints, Knight and Viner,⁴⁴ a tradition more flexible than the Hayekian version dominant at LSE.⁴⁵ Because of this tradition, Friedman argued, Chicago economists were less vulnerable to the lure of Keynesian ideas. There is considerable evidence, however, that Friedman's restatement should rather be regarded as a new interpretation of the quantity theory, albeit one following on closely from the Chicago tradition. In addition, there is evidence that later versions of this Chicago tradition owed something to the Keynesian theory of liquidity preference.⁴⁶

Whether a restatement of an earlier tradition, or an original interpretation of the quantity theory, Friedman's article set the pattern for the development of the quantity theory for the next decade. Friedman himself undertook empirical investigations into the two assertions listed above, research culminating in *A Monetary History of the United States, 1861–1960* (1963, with Schwarz).⁴⁷ In these studies he reached the conclusion that the supply of money had been to a substantial extent independent of demand, and that money had a strong influence on the economy. A particularly important conclusion was that there was a long and variable lag between monetary changes and their effects on the economy. This link, however, was strongest for large changes (deep depression cycles and substantial inflations): if changes in the money supply were only moderate then other factors were also important.⁴⁸ Though Friedman found a strong link operating from money to prices and income, he was able to defend the policy he had long since advocated (in *A Program for Monetary Stability*, 1946): that the object of monetary policy should be to expand the money supply at a steady rate.⁴⁹ Limitations of knowledge concerning the other factors affecting the economy, factors which were important in mild cycles, imposed definite limits to the possibilities for fine tuning the economy

through monetary policy. The most that could be achieved was to prevent money from being a major source of disturbance.

Of particular interest is an episode which arose out of an attempt by Friedman and Meiselman (1963) to find a way of testing the quantity theory against the Keynesian theory. The test they proposed was to compare the “money multiplier” (relating income to the quantity of money) with the Keynesian investment multiplier. If the former was more stable, as Friedman and Meiselman contended, this was evidence for the quantity theory. This challenge was taken up, in particular by Ando and Modigliani (1965). Despite a whole issue of the *American Economic Review* being devoted to the question, the debate was inconclusive: it became agreed that simple models of this type were incapable of discriminating between the two theories.⁵⁰

Throughout these discussions of the quantity theory Friedman had never provided an alternative to the Keynesian theory of how money was supposed to affect the economy – a theory of the monetarist “transmission mechanism”. The overwhelming emphasis in his work was on empirical investigations based on a loosely specified theory. This situation changed in 1970–1971 when Friedman published two papers, later combined under the title “A theoretical framework for monetary analysis”.⁵¹ Here Friedman proposed a common framework in which Keynesianism and monetarism could be compared. This common framework was the Hicksian IS–LM model. As there were three variables to be determined, but only two equations (the IS and LM relations) there was a “missing equation”. The simple quantity theory took output as given; the simple Keynesian theory added a fixed price level. Friedman added a theory of how changes in nominal income were broken down into changes in output and the price level.⁵²

This discussion of Friedman’s theoretical framework, however, was in many ways a hangover from the 1960s, for in the 1970s attention shifted, mainly because of the expectations augmented Phillips curve and the natural rate hypothesis. Attention shifted towards the dynamics of inflation and unemployment, and to the question of how expectations were formed. The arguments for monetarism came to centre on the dangers of accelerating inflation if the authorities tried to control unemployment. This argument and Friedman’s earlier arguments could be combined to produce a case for aiming at a steady growth rate of the money supply: the government should try to control only what it is capable of controlling (the money supply, not unemployment), and because the lags involved made monetary policy hard to use in a discretionary manner, a fixed monetary rule should be pursued.

Tobin’s portfolio balance approach

An alternative approach to the theory of monetary policy was adopted by Tobin. Like Keynes and Friedman, Tobin viewed money as an asset, but rather than distinguish a unique asset, or set of assets, to be called money, Tobin considered a variety of assets ranging from currency at one end to physical capital at the other.⁵³ He constructed a series of models of the

financial sector in which a market was assumed to exist for each asset – currency and various types of bank deposit being considered separately. In these models the crucial variable was the valuation ratio, Tobin's "q", which is the ratio of the market valuation of the capital stock to its value at replacement cost. It is this variable which measures the incentive to invest. The valuation of the capital stock, and hence "q", depends on supply and demand, with demand depending on the yield on capital relative to the yields available on other assets.

The significance of this approach is twofold. It stresses the existence of a range of assets, showing that the basic IS–LM model can still be used, even to represent an economy with a complicated financial sector. In Tobin's models the LM curve, instead of describing equilibrium simply in the money market, results from the equilibrium of the whole financial sector.⁵⁴ The approach recognizes the problems involved in defining money.⁵⁵ It may be possible to group together a collection of assets and to call them money, but there is no need to do this. Monetary economics can be done satisfactorily in terms of a continuous range of assets.

26.3 RATIONAL EXPECTATIONS AND THE NEW CLASSICAL MACROECONOMICS

Rational expectations

In the mid 1970s macroeconomics was transformed by the introduction and systematic use of the concept of "rational expectations", the chief architects of this being Lucas, Sargent, Wallace and Barro.⁵⁶ Though the concept of rational expectations was not new, having been developed in a different context by Muth in the early 1960s,⁵⁷ it was only in the 1970s that economists saw the implications of the concept for macroeconomics and developed them.

The basic idea underlying rational expectations is that people learn, and that they will form their expectations on the basis of all the information they have available. Their behaviour in acquiring and using information will be governed by utility maximization: if it is profitable to acquire new information, or to use the information they have in a more efficient way then people will do so. To analyse such a process of learning is, however, immensely complicated, an alternative to which is to analyse an equilibrium in the process of learning: to analyse a situation where the process of learning has come to an end in the sense that people have no incentive to change the method by which they form their expectations. The simplest such situation to analyse is the situation where people have learnt everything there is to be learnt: in other words, a situation where everything that is predictable is predicted correctly. If there is no uncertainty this amounts to assuming that expectations are correct. If there is uncertainty it implies that the errors people make are completely random and unpredictable.

Why did rational expectations catch on so fast in the 1970s? Two reasons immediately suggest themselves. The first is that work on expectations-augmented Phillips curves showed that the formation of inflationary expectations was crucial in determining how an economy worked, in particular how it responded to monetary and fiscal policy. There was, however, no satisfactory theory of expectations. Expectations were explained by economists in terms of various ad hoc rules, such as “adaptive expectations”,⁵⁸ all of which implied that people were failing to use the information available to them. There was a clear need for a better theory of expectations. The second reason is that the concept represents the application of individual maximizing behaviour, an assumption that was being progressively applied to other areas of economic theory, to the formation of expectations.

The new classical macroeconomics

The new classical macroeconomics combines the hypothesis of rational expectations with the natural rate hypothesis. The implications of this were dramatic, as was shown in a series of papers in the mid 1970s.⁵⁹ By the mid 1970s the Phelps–Friedman expectations-augmented Phillips curve was becoming generally accepted, but there was still controversy as to whether the long run Phillips curve was completely vertical, this hinging on whether expectations responded fully to changes in the inflation rate. The introduction of rational expectations swept all this away. If expectations are rational, they must in the long run be correct on average: anything different would imply that people were ignoring easily available information. The new classical argument, however, went much further than this. The Phelps–Friedman argument had admitted that stabilization policy was not in principle impossible. It was the practical objections which were decisive. The new classical argument, on the other hand, was that it was impossible, unless the government had more information about the economy than did the private sector, for government policy, however efficiently conducted, to have any systematic effect on output. Policy could affect unemployment only by causing errors in expectations, and by the assumption of rational expectations these errors must be random and unpredictable. There was no usable trade-off, even in the short run.

Of fundamental importance to the new classical view of economic policy is what is known as the “Lucas critique” of stabilization policy,⁶⁰ for this was a critique of the whole approach to macroeconomic policy prevalent since 1945. The traditional approach to economic policy is to estimate a model of the economy comprising a series of equations describing how the private sector responds to changes in exogenous variables. This model can then be used to explore the implications of various policy proposals with a view to choosing the best of these, according to whatever criterion the policymakers consider appropriate. Lucas’ argument is that this approach is misconceived, for the behaviour of the private sector (and hence the initial macroeconomic model) depends on what the public believe policymakers to

be doing. If policy changes, so will the model. Macroeconomic policy must be viewed as a process where the behaviour of the private sector depends on the policy being pursued by the government.

One of the attractions of the new classical macroeconomics is its ability to encompass a wide range of issues within a coherent framework. The central and overriding theme is individual rationality, the hypothesis of utility maximization being taken to its extreme. All else is subordinate to this axiom: rational expectations result from its application to the formation of expectations; continuous market clearing from its application to market behaviour (if markets do not clear people are ignoring potentially profitable opportunities for trade). It is taken as a methodological datum that explanation must be in terms of maximizing behaviour, not arbitrarily specified functions.⁶¹

Responses to the new classical macroeconomics

The ideas discussed in the previous two sections are very recent and it is much too early to assess their long term fate. Several points can, however be made. (1) Rational expectations (as distinct from the new classical macroeconomics) has established itself as an invaluable theoretical tool. It has become widely accepted, even amongst opponents of the new classical macroeconomics, that, in the absence of evidence on why expectations should differ from rational expectations, it is important at least to examine the solutions of models under rational expectations. (2) The new classical macroeconomics has unleashed a wealth of theoretical and empirical work, especially in the US. Furthermore it is still developing, as evidenced by the recent revival of the real bills doctrine.⁶² (3) Questions of supply have been given a more prominent place, this being explicable in terms not only of theoretical developments, but also in terms of circumstances. (4) The Keynesian–Monetarist distinction has become an irrelevance in analysing attitudes to macroeconomics. The views now held by many economists commonly regarded as Keynesian (e.g. Tobin, Modigliani, Solow) are, though they remain critical of both Friedman and the new classical economics, very different from the views held by Keynesians twenty years ago.⁶³ (5) Economists have started to analyse models incorporating rational expectations, but in which the new classical conclusions do not follow.⁶⁴ Though such work is in its infancy it is clear that the new classical economics has stimulated economists to ask new questions, and to approach old questions in new ways.

26.4 THE BUSINESS CYCLE

Multiplier–accelerator models

The *General Theory* had an enormous impact on business cycle theory. As early as 1936 Harrod's *The Trade Cycle* used a multiplier–accelerator model,

though it was Samuelson's (1939b) version which became the standard exposition. The attraction of this approach compared with earlier ones was that a cycle emerged naturally from the interaction of two simple relations. The accelerator explained why, for example, a fall in the *growth rate* of consumption would produce a fall in the level of investment. This in turn would, via the multiplier, lower consumption, producing a turning point in output. This approach, together with variations such as Metzler's inventory cycle (1941), dominated business cycle theory in the 1940s and early 1950s.

The Keynesian origins of this approach to the business cycle are clear, but it is important not to neglect other influences. (i) There was the accelerator itself, due in particular to Aftalion and Clark.⁶⁵ (ii) Furthermore, if the interaction of the multiplier and the accelerator were to produce cycles, lags were needed, these not being provided by Keynes, who used a static model even when this was inappropriate, but by others. Robertson (1933, 1936) postulated a lag between demand and income; Lundberg (1937) one between output and income. These changes brought the theory closer in some respects to earlier theories, such as the Swedish attempts at period analysis. (iii) Having set up a model with lags it had to be analysed, and here the application, following Frisch (1933), of mathematical methods was crucial. A critical property of the multiplier–accelerator model was, as Samuelson showed, that it would generally produce either cycles which faded away, or cycles of ever-increasing amplitude. The “realistic” case of fairly regular cycles would occur only by chance. (iv) Finally, because the multiplier–accelerator model could not pretend to be a complete model of the cycle, other factors had to be brought in. For example, Hicks' *Contribution to the Theory of the Trade Cycle* (1950), described by one authority as “the most elegant and most carefully elaborated specimen of a great variety of similar systems”,⁶⁶ used a model where an unstable multiplier–accelerator process operated between a ceiling imposed by full capacity, and a floor imposed by a minimum below which investment could not fall.

Before going on to consider alternative approaches to the cycle, we need to say something about the thesis, closely tied up with views of the cycle, of “secular stagnation”. Although this idea was present in the *General Theory*,⁶⁷ Keynes viewing the prospect favourably, it is a view associated, above all, with Hansen.⁶⁸ Hansen, the foremost American apostle of Keynesian ideas in the 1940s, arrived at this conclusion through combining the Keynesian theory with arguments taken from earlier business cycle theories, a mixture described by one critic as “a strange amalgam of Keynes, Schumpeter and Spiethoff”.⁶⁹ The basic argument was that as the US economy matured, and the growth rate of population and resources fell, opportunities for investment would be exhausted. Autonomous investment would fall. At the same time there was a tendency for savings to rise with prosperity, reducing the multiplier. Thus government intervention, such as deficit spending, would be necessary to sustain growth and to prevent stagnation.

Econometric models

The Hicksian approach to the cycle was to assume a potentially explosive multiplier–accelerator process, constrained by a floor and a ceiling. The alternative approach was to use a stable multiplier–accelerator process, one in which cycles would, if left undisturbed, die away, and to postulate that it was kept going by a series of random shocks. Econometric models were important here for two reasons. The first was that evidence from such models was that they seemed to be inherently stable: in the absence of shocks, fluctuations would die away. This cast doubt on the Hicksian approach. The second, and more fundamental reason, was that this type of model requires different techniques from those required by the Hicksian approach, methods more akin to those required in econometrics. Particularly fruitful are simulation methods, an important such study being Adelman and Adelman (1959). This study took the Klein–Goldberger (1955) model of the US economy and solved it under various assumptions. They found that the model itself did not generate cycles: it settled down rapidly after a disturbance. But when a suitable series of random shocks was introduced, the model produced cyclical fluctuations considered “remarkably similar” to those described by the National Bureau.⁷⁰

These results are interesting because they illustrate two things. (i) Progress in economic theory is dependent on the availability of appropriate techniques: in this instance the development of simulation methods, this in turn being dependent on the availability of computers. (ii) The availability of techniques may influence the way in which the economy is perceived. Thus the use of a simple second-order difference equation to represent cyclical fluctuations suggested a clear-cut difference between the Hicksian use of the multiplier–accelerator model, according to which cycles result from the structure of the economy, and the Frisch–Adelman view, according to which they result from the way in which the economic system responds to exogenous shocks.

Money and the business cycle

Though not alone in arguing that monetary factors had, under Keynesian influence, been unduly neglected,⁷¹ responsibility for arousing interest in the monetary aspects of business cycles rests above all with Friedman. In the tradition of Mitchell and the National Bureau, under the aegis of which much of his research was undertaken, Friedman’s approach was primarily empirical, being concerned with the behaviour of money, income and the velocity of circulation over the cycle.⁷² Typical is “Money and business cycles” (1963b), where extensive discussion of empirical evidence is followed by what Friedman and Schwarz describe as “a tentative sketch” of mechanisms linking monetary changes to the cycle. Aside from his contributions to the quantity theory more generally, Friedman’s main contribution was the use of permanent income to explain cyclical variations in the velocity of circulation.⁷³

Despite Friedman's advocacy, however, monetary theories of the cycle failed to catch on. In part this was due to Friedman's approach, with its attempt to demonstrate the importance of money through examining evidence on the timing of cycles in money income and velocity. As Tobin (1970) pointed out with some carefully worked out counter-examples, empirical evidence on timing may be able to refute the theory that monetary factors are responsible for cycles, but it cannot possibly establish it. Furthermore, attempts to develop a monetary theory to pose against the Keynesian theories could hardly proceed until a suitable theory of the relationship between money, output and prices had been developed. Such a theory was not available.

After a period of comparative neglect in the 1960s, when interest in dynamics was focused instead on growth, interest in the business cycle, and with it interest in monetary theories of the cycle, was revived by the work of Lucas in the mid 1970s.⁷⁴ In the same way, however, that the new classical macroeconomics constituted a radical departure from the more traditional monetarism associated with Friedman, so the new classical theory of the business cycle was radically different from that of Friedman and Schwarz. The object was to provide an "equilibrium" theory of the cycle: one in which all markets were assumed to be in equilibrium continuously, and in which expectations were rational. The source of fluctuations was found in errors in inflationary expectations, these in turn resulting from unanticipated changes in the growth rate of the money supply. The interpretation of the resulting cycles is, however, very different from that of fluctuations in Keynesian theory. Unemployment arises in equilibrium theories of the cycle because workers choose to work less because of mistakes they make in evaluating current and future real wage rates. Thinking that the current real wage rate is unduly low, for example, workers choose to take leisure now rather than in the future.

The fundamental axioms of the new classical macroeconomics, however, are insufficient in themselves to generate a cycle. As expounded so far, the theory can only explain random fluctuations in output, whereas business cycles are not random – the essence of the business cycle is that fluctuations in output follow a pattern. To explain this pattern lags have to be introduced into the model. Whilst there is nothing inherently wrong with this, it is important to notice that it is not the equilibrium elements of the theory that produces cycles: it is just as much the factors introduced from outside – the lags, about which macroeconomic theory has little to say. The theory looks very close to the econometric theories discussed above: the cycle results from the interaction of the economic system with a system of exogenous shocks.

26.5 CONCLUSIONS

It would be natural to think of the conflict between Keynesianism and monetarism as the main feature of post-war macroeconomics, and to a great

extent this is correct. It is arguable, however, that the dominant theme, from Hicks's work on Keynes and the "classics" to Lucas's work on rational expectations, has been the attempt to establish a macroeconomic theory based on secure microeconomic foundations. This has been a crucial factor, even in areas (such as the theory of the consumption function, or the theory of inflation and unemployment) where econometric work has been undertaken on a large scale. It is, on the whole, developments in theory, not the results of empirical work, which change the way economists conceive the economy. Thus, though interest in a new line of inquiry has sometimes been prompted by the discovery of a new empirical regularity (e.g. the Phillips curve), it has been more common for a new departure in macroeconomics to be the result of a new theoretical idea. Important examples are Hicks's interpretation of Keynes in terms of a general equilibrium model, the "general disequilibrium" theory based on the work of Clower and Leijonhufvud, and the new classical macroeconomics.