Nobel Lecture: Monetary Neutrality

Robert E. Lucas, Jr.

University of Chicago

I. Introduction

The work for which I have received the Nobel Prize was part of an effort to understand how changes in the conduct of monetary policy can influence inflation, employment, and production. So much thought has been devoted to this question and so much evidence is available that one might reasonably assume that it had been solved long ago. But this is not the case: It had not been solved in the 1970s when I began my work on it, and even now this question has not been given anything like a fully satisfactory answer. In this lecture I shall try to clarify what it is about the problem of bringing available evidence to bear on the assessment of different monetary policies that makes it so difficult and to review the progress that has been made toward solving it in the last two decades.

From the beginnings of modern monetary theory, in David Hume's marvelous essays of 1752, *Of Money* and *Of Interest,* conclusions about the effect of changes in money have seemed to depend critically on the way in which the change is effected. In formulating the doctrine that we now call the quantity theory of money, Hume stressed the units-change aspect of changes in the money stock and the irrelevance of such changes to the behavior of rational people.

> It is indeed evident that money is nothing but the representation of labour and commodities, and serves only as a method of rating or estimating them. Where coin is in

I thank Nancy Stokey for invaluable discussion and criticism. I am also very grateful for comments from William Brock, John Cochrane, Milton Friedman, Lars Hansen, Anil Kashyap, Randall Kroszner, Bennett McCallum, Casey Mulligan, Sherwin Rosen, Allen Sanderson, Thomas Sargent, John Taylor, Neil Wallace, Warren Weber, and Jörgen Weibull.

[*Journal of Political Economy,* 1996, vol. 104, no. 4]

> greater plenty, as a greater quantity of it is required to represent the same quantity of goods, it can have no effect, either good or bad . . . any more than it would make an alteration on a merchant's books, if, instead of the Arabian method of notation, which requires few characters, he should make use of the Roman, which requires a great many. [*Of Money*, p. 32][1]

Hume returns to this idea that changes in the quantity of money are just units changes in *Of Interest:*

> Were all the gold in England annihilated at once, and one and twenty shillings substituted in the place of every guinea, would money be more plentiful or interest lower? No surely: We should only use silver instead of gold. Were gold rendered as common as silver, and silver as common as copper, would money be more plentiful or interest lower? We may assuredly give the same answer. Our shillings would then be yellow, and our halfpence white, and we should have no guineas. No other difference would ever be observed, no alteration on commerce, manufactures, navigation, or interest, unless we imagine that the colour of the metal is of any consequence. [p. 47]

These are two of Hume's statements of what we now call the quantity theory of money: the doctrine that changes in the number of units of money in circulation will have proportional effects on all prices that are stated in money terms, and no effect at all on anything real, on how much people work or on the goods they produce or consume. Notice, though, that there is something a little magical about the way in which changes in money come about in Hume's examples. All the gold in England gets "annihilated." Elsewhere he asks us to "suppose that, by miracle, every man in Great Britain should have five pounds slipt into his pocket in one night" (p. 51). Money changes in reality do not occur by such means. Is this just a matter of exposition, or should we be concerned about it? This turns out to be a crucial question. In fact, Hume writes as follows:

> When any quantity of money is imported into a nation, it is not at first dispersed into many hands but is confined to the coffers of a few persons, who immediately seek to employ it to advantage. Here are a set of manufacturers or merchants, we shall suppose, who have received returns of gold and

[1] All page references to Hume's essays are taken from Hume (1970). I have left the spelling as in the original and modernized the punctuation.

> silver for goods which they sent to Cadiz. They are thereby enabled to employ more workmen than formerly, who never dream of demanding higher wages, but are glad of employment from such good paymasters. . . . [The artisan] carries his money to market, where he finds every thing at the same price as formerly, but returns with greater quantity and of better kinds for the use of his family. The farmer and gardener, finding that all their commodities are taken off, apply themselves with alacrity to raising more. . . . It is easy to trace the money in its progress through the whole commonwealth, where we shall find that it must first quicken the diligence of every individual before it encrease the price of labour. [p. 38]

Symmetrically, Hume believes that a monetary contraction could induce depression:

> There is always an interval before matters be adjusted to their new situation, and this interval is as pernicious to industry when gold and silver are diminishing as it is advantageous when these metals are encreasing. The workman has not the same employment from the manufacturer and merchant, though he pays the same price for everything in the market. The farmer cannot dispose of his corn and cattle, though he must pay the same rent to his landlord. The poverty, and beggary, and sloth which must ensue are easily foreseen. [p. 40]

Hume makes it clear that he does not view his opinions about the initial effects of monetary expansions as major qualifications to the quantity theory, to his view that "it is of no manner of consequence, with regard to the domestic happiness of a state, whether money be in a greater or less quantity" (p. 39). Perhaps he simply did not see that the irrelevance of units changes from which he deduces the long-run neutrality of money has similar implications for the initial reaction to money changes as well. Why, for example, does an early recipient of the new money "find every thing at the same price as formerly"? If everyone understands that prices will ultimately increase in proportion to the increase in money, what force stops this from happening right away? Are people committed, perhaps even contractually, to continue to offer goods at the old prices for a time? If so, Hume does not mention it. Are sellers ignorant of the fact that money has increased and a general inflation is inevitable? But Hume claims that the real consequences of money changes are "easy to trace" and "easily foreseen." If so, why do these consequences occur at all?

These questions do not involve mere matters of detail. Hume has deduced the quantity theory of money by purely theoretical reasoning from "that principle of reason" that people act rationally and that this fact is reflected in market-determined quantities and prices. Consistency surely requires at least an attempt to apply these same principles to the analysis of the *initial* effects of a monetary expansion or contraction. I think the fact is that this is just too difficult a problem for an economist equipped with only verbal methods, even someone of Hume's remarkable powers.

This tension between two incompatible ideas—that changes in money are neutral units changes and that they induce movements in employment and production in the same direction—has been at the center of monetary theory at least since Hume wrote. Though it has not, in my opinion, been fully resolved, important progress has been made on at least two dimensions. The first is a purely theoretical question: Under what assumptions and for what kinds of changes can we expect monetary changes to be *neutral*? (I take this terminology from Don Patinkin's *Money, Interest, and Prices* [1965], the book that introduced so many economists of my cohort to these theoretical issues.) The theoretical equipment we have for sharpening and addressing such questions has been vastly improved since Hume's day, and I shall draw on these improvements below. Of at least equal importance, an enormous amount of evidence on money, prices, and production has been accumulated over the past two centuries, and much fruitful thought has been applied to issues of measurement. In the next section, I shall examine some of this evidence.

II. Evidence

It is hard to tell from the essays what evidence Hume actually had in front of him. Certainly he wrote before systematic data on money supplies were collected anywhere in the world, before the invention of price indexes, and long before the invention of national income and product accounting. His development of the quantity theory was based largely on purely theoretical reasoning, though tested informally against his vast historical knowledge, and his belief in short-run correlations between changes in money and changes in production was apparently based mainly on his everyday knowledge. (He cites one Mons. du Tot for the assertion that "in the last year of Louis XIV, money was raised three-sevenths but prices augmented only one" [p. 39]. In a footnote he characterizes his source as "an author of reputation," but feels obliged to "confess that the facts which he

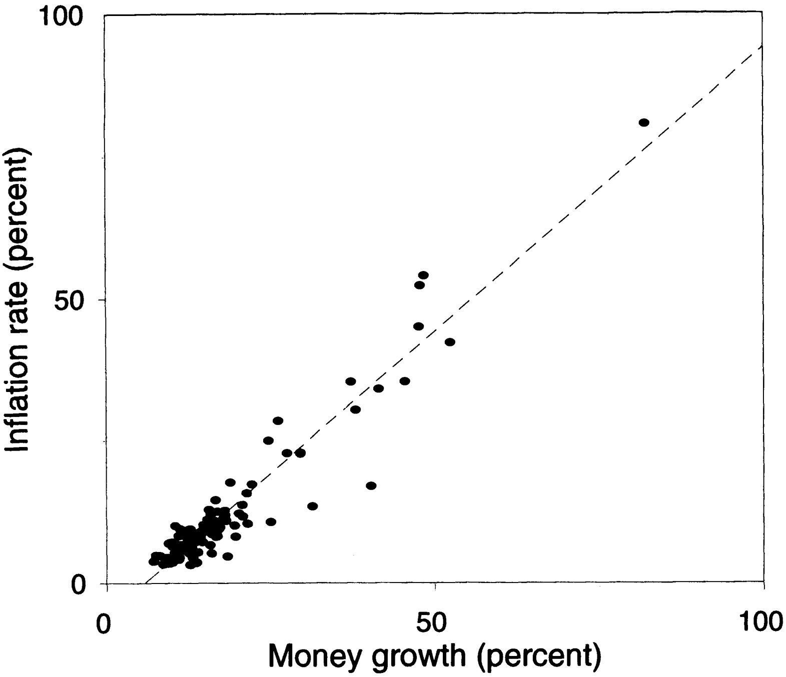

FIG. 1

advances on other occasions are often so suspicious as to make his authority less in this matter." Even in the eighteenth century, it seems, there were tensions between theorists and econometricians!)

The central predictions of the quantity theory are that, in the long run, money growth should be neutral in its effects on the growth rate of production and should affect the inflation rate on a one-for-one basis. The modifier "long run" is not free of ambiguity, but by any definition the use of data that are heavily averaged over time should isolate only long-run effects. Figure 1, taken from McCandless and Weber (1995), plots 30-year (1960–90) average annual inflation rates against average annual growth rates of M2 over the same 30-year period, for a total of 110 countries. One can see that the points lie roughly on the 45-degree line, as predicted by the quantity theory. The simple correlation between inflation and money growth is .95. The monetary aggregate used in constructing figure 1 is M2, but nothing important depends on this choice. McCandless and Weber report a simple correlation of .96 if M1 is used and .92 if M0 (the monetary base) is used. They also report correlations for subsets of their 110-country data set: .96 (with M2) with only OECD countries and .99 with 14 Latin American countries.

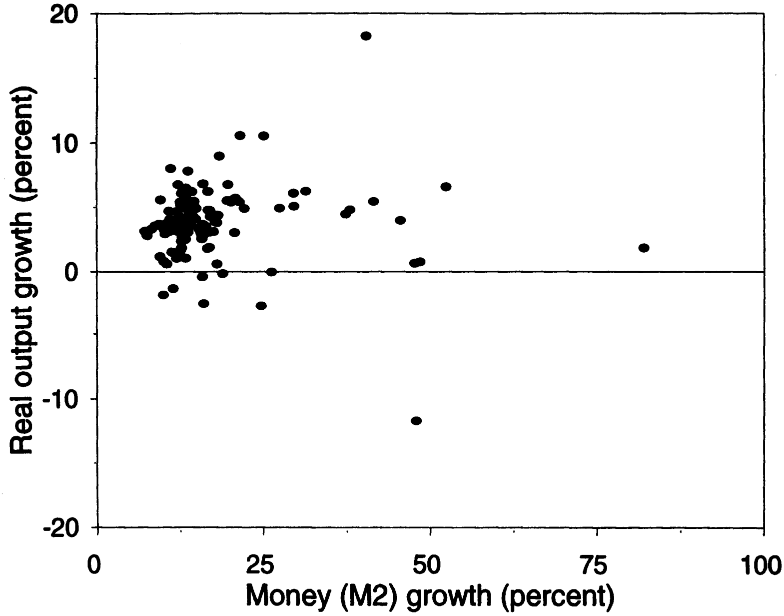

FIG. 2

It is clear from these data (and from the many other studies that have reached similar conclusions) that the applicability of the quantity theory of money is not limited to currency reforms and magical thought experiments. It applies, with remarkable success, to comovements in money and prices generated in complicated, real-world circumstances. Indeed, how many specific economic theories can claim empirical success at the level exhibited in figure 1? Central bankers and even some monetary economists talk knowledgeably of using high interest rates to control inflation, but I know of no evidence from even one economy linking these variables in a useful way, let alone evidence as sharp as that displayed in figure 1. The kind of monetary neutrality shown in this figure needs to be a central feature of any monetary or macroeconomic theory that claims empirical seriousness.

McCandless and Weber also provide evidence on correlations between money growth and growth in real output, averaged over the 1960–90 period. Figure 2 is their plot for the full 100-country data set from the International Monetary Fund. Evidently, there is no relation between these 30-year averages.[2] For examining short-term

[2] It must be said that the evidence of long-run links between money growth and output growth is more mixed than one would infer from fig. 2. McCandless and Weber find a weak positive relation for the OECD economies. Both positive and negative correlations have been found by other investigators on other data sets.

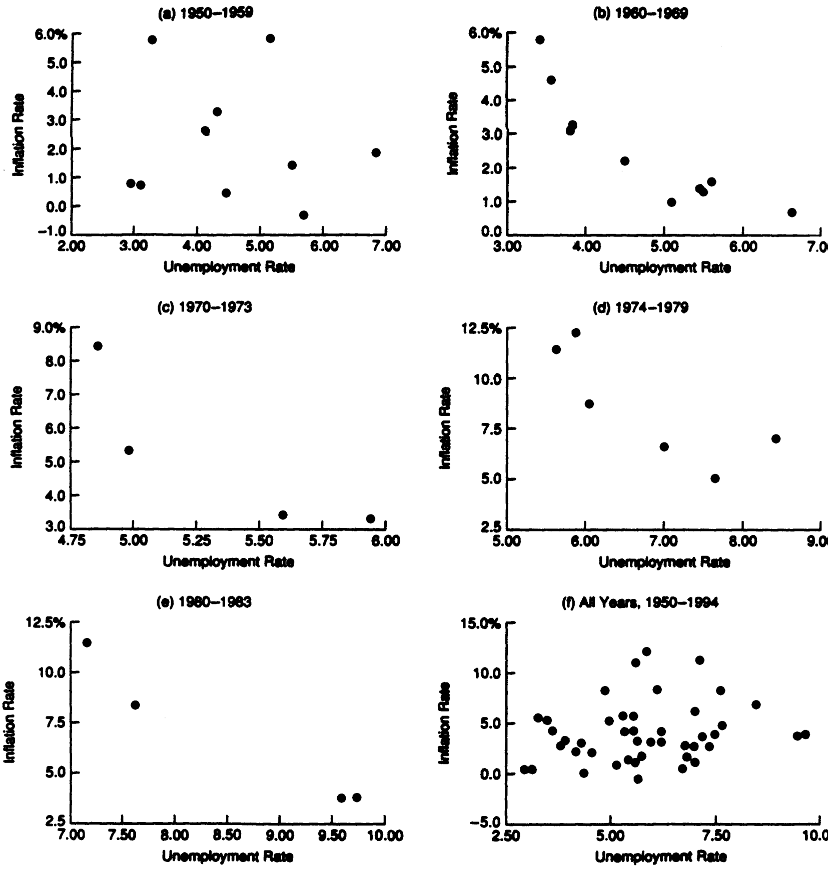

FIG. 3.—Inflation and unemployment in the United States

trade-offs, of course, one does not want to use such time-averaged data. Figure 3, taken from Stockman (1996), provides six plots of annual inflation rates against unemployment rates for various subperiods of the years 1950–94, for the United States. Panel *f* plots the Phillips curve (after A. W. Phillips [1958]) for the entire period. In this panel, the two variables appear to be completely unrelated. On the other hand, the five panels for subperiods (or at least for the subperiods since 1960) seem to show a clear, negative relation. But then look at the axes in these six panels! In order to see inflation and unemployment as lying on a negatively sloped curve, one needs to keep shifting the curve.

Evidence on trade-offs is also marshaled, though in a very different way, in Friedman and Schwartz's (1963) monograph *A Monetary History of the United States.* These authors show that every major depression in the United States over the period 1867–1960 was associated with a large contraction in the money supply and that every large contraction was associated with a depression. These observations are correlations of a sort, too, but they gain force from the size of the largest contractions. In a period such as the post–World War II years in the United States, real output fluctuations are modest enough to be attributable, possibly, to real sources. There is no need to appeal to money shocks to account for these movements. But an event such as the Great Depression of 1929–33 is far beyond anything that can be attributed to shocks to tastes and technology. One needs some other possibilities. Monetary contractions are attractive as the key shocks in the 1929–33 years, and in other severe depressions, because there do not seem to be any other candidates.

Sargent (1986) also examines large, sudden reductions in rates of money growth (though not reductions in the *levels* of money stocks). In his case, these are the monetary and fiscal reforms that ended four of the post–World War I European hyperinflations. These dramatic reductions in growth rates of the money supply dwarf anything in Friedman and Schwartz or in the postwar data used by McCandless and Weber. Yet as Sargent shows, they were not associated with output reductions that were large by historical standards, or possibly by any depressions at all. Sargent goes on to demonstrate the likelihood that these reductions in money growth rates were well anticipated by the people they affected and, because of visible and suitable fiscal reforms, were expected by them to be sustained.

In summary, the prediction that prices respond proportionally to changes in money in the long run, deduced by Hume in 1752 (and by many other theorists, by many different routes, since), has received ample—I would say decisive—confirmation, in data from many times and places. The observation that money changes induce output changes in the same direction receives confirmation in some data sets but is hard to see in others. Large-scale reductions in money growth can be associated with large-scale depressions or, if carried out in the form of a credible reform, with no depression at all.

III. Theoretical Responses

Hume was able to theorize rigorously, and, as we have seen, with great empirical success, about comparisons of long-run average behavior across economies with different average rates of money growth. For short-run purposes, on the other hand, he was obliged to rely on much looser reasoning and rough empirical generalizations. As economic theory evolved in the last century and most of this one, the double standard that characterized Hume's argument was perpetuated. The quantity-theoretic "neutrality theorems" were stated with increasing precision and worked through rigorously, using the latest equipment of static general equilibrium theory. The dynamics had a kind of patched-in quality, fitting the facts, but only in a manner that suggests they could equally well fit *any* facts. Patinkin interprets all of monetary theory from Wicksell's *Interest and Prices* (1898) through his own *Money, Interest, and Prices* as concerned with processes of adjustment between one quantity-theoretic equilibrium position and another, conceived as outside the framework of general equilibrium theory in a way that seems to me very much at the level of Hume's analysis. The passages on dynamics that I cited from Hume in Section I could be slipped into Keynes's *Treatise on Money* (1930) or Hayek's *Monetary Theory and the Trade Cycle* (1933) without inducing any sense of anachronism.

Yet all these theorists *want* to think in general equilibrium terms, to think of people as maximizing over time, as substituting intertemporally. They resort to disequilibrium dynamics only because the analytical equipment available to them offers no alternative. Even in Hume's scenario, the motives and expectations of economic actors during the transition are described, even rationalized: The adjustment to a new equilibrium is not seen as a purely mechanical *tâtonnement* process, the character of which is determined by forces apart from the producers and consumers of the system. Certainly Wicksell and, I would say, Patinkin, too, are trying to think through the way the dynamic adjustment process appears to people as it occurs, to see the actions people take as rational responses to their situations. Though the theoretical formalism on which they draw involves a static equilibrium combined with a mechanical process to describe dynamic adjustments, their verbal descriptions of periods of transition, like Hume's before them, show that they are in fact thinking of people solving intertemporal decision problems.

The intelligence of these attempts to deal theoretically with the real effects of changes in money is still impressive to the modern reader, but serves only to underscore the futility of attempting to talk through hard dynamic problems without any of the equipment of modern mathematical economics. Hayek and Keynes and their contemporaries were willing to make assumptions and to set out something like a model, but they were simply not able to work out the predictions of their own theories.

The depression of the 1930s shifted attention away from the subtle problems of monetary neutrality and toward the potential of monetary policy for short-run stimulus. Keynes's *General Theory* (1936) was one product of this change of focus. Another was Tinbergen's (1939) development of an explicit statistical model of the U.S. economy.

Tinbergen's model and its immediate successors made little or no contact with earlier traditions in monetary theory; but in the atmosphere of the 1930s and 1940s this was perhaps an advantage, and it fit well with the revolutionary rhetoric of Keynes's book. Economic theory aside, the macroeconometric models that evolved from Tinbergen's work had two important advantages over all earlier macroeconomic theory. They were mathematically explicit and so could be estimated from and tested against data in a much more disciplined way than earlier theories could. Moreover, they could be simulated to yield quantitative answers to policy questions. It was these features that excited younger researchers and had such a dramatic influence on future developments in the field.

By the 1960s, then, two very different styles of macroeconomic theory, both claiming the title of Keynesian economics, coexisted. There were attempts at a unified monetary and value theory, like Patinkin's, based on extensions of static general equilibrium theory to accommodate money, combined with some kind of *tâtonnement* process to provide some dynamics. These theories were developed with great attention to earlier monetary theory and to developments in economic theory more generally, but they lacked the operational character of the macroeconometric models: No one could tell exactly what their predictions were or what normative implications they carried. On the other hand, there were macroeconometric models that could be fit to data and simulated to yield quantitative answers to policy questions but whose relation to microeconomic theory and classical monetary theory was unclear.

Virtually no one at the time regarded this situation as healthy. Everyone paid lip service to the idea of unification of micro- and macroeconomics or of discovering the microeconomic "foundations" of macroeconomic theories, and a vast amount of creative and valuable economics, focused on intertemporal decision problems, was inspired by this goal. The work of Modigliani and Brumberg (1954) and Friedman (1957) on consumption; Eisner and Strotz (1963) and Jorgenson (1963) on investment in physical capital; Becker (1962) and Ben-Porath (1967) on investment in human capital; and Baumol (1952), Tobin (1956), Brunner and Meltzer (1963), and Meltzer (1963) on money demand all contributed. Mathematically inclined economists who entered the profession in the 1960s were drawn to methods for studying intertemporal decision problems, the calculus of variations, the theory of optimal control, and Bellman's (1957) dynamic programming. Work on optimal growth by Uzawa (1964) and Cass (1965) was followed by applications of similar methods to a variety of problems in all subfields of economics. In these applications, the dynamics were an integral part of the theory, not tacked-on *tâtonnement* processes. When Leonard Rapping and I began our collaboration on the study of labor markets over the business cycle (Lucas and Rapping 1969), we thought of these studies of individual intertemporal decision problems as models of what we wanted to do.[3]

The prevailing strategy for macroeconomic modeling in the early 1960s held that the individual or sectoral models arising out of this intertemporal theorizing could then simply be combined in a single model, the way Keynes and Tinbergen and their successors assembled a consumption function, an investment function, and so on into a model of an entire economy. But models of individual decisions over time necessarily involve expected, future prices. Some microeconomic analyses treated these prices as known; others imputed adaptive forecasting rules to maximizing firms and households. However it was done, though, the "church supper" models assembled from such individual components implied behavior of *actual* equilibrium prices and incomes that bore no relation to, and were in general grossly inconsistent with, the price expectations that the theory imputed to individual agents.

As intertemporal elements and expectations came to play an increasingly explicit and important role, this modeling inconsistency became more and more glaring. John Muth's (1961) "Rational Expectations and the Theory of Price Movements" focused on this inconsistency and showed how it could be removed by taking into account the influences of prices, including *future* prices, on quantities and *simultaneously* the effects of quantities on equilibrium prices.[4] The principle of rational expectations he proposed thus forces the modeler toward a market equilibrium point of view, although it took some time before a style of thinking that recognized this fact had a major effect on macroeconomic modeling.

Other considerations reinforced a move in the same direction. In the late 1960s, Friedman (1968) and Phelps (1968) saw, by thinking through the issue at a general equilibrium level, that there could be no long-run Phillips curve trade-off between inflation and real output. But such long-run trade-offs were implicit in all the macroeconometric models of the day, and the econometric methods that were in standard use at that time seemed to reject the Friedman-Phelps natural rate hypothesis. This conflict led to a rethinking of the theoretical basis of these statistical tests, and the discovery of serious difficulties with them. Sargent (1971) and Lucas (1972*a*, 1976) showed that the

[3] See Phelps et al. (1970) for several similarly motivated studies.

[4] Eugene Fama's (1965) theory of efficient markets was another direct application of economic reasoning to the behavior of *equilibrium* prices, in a setting in which stochastic shocks were an intrinsic part of the economic model.

conventional rejections of the natural rate hypothesis depended critically on *irrational* expectations or, to put the same point backward, that if one assumed rational expectations these tests settled nothing. It seemed clear that it was necessary to put macroeconomics on a general equilibrium basis that incorporated rational expectations.

IV. General Equilibrium Macroeconomics

By the 1960s, two closely related general equilibrium frameworks were in fact already available for thinking about economic dynamics. One was the mathematical model of general equilibrium, developed by Hicks (1939), Arrow (1951), McKenzie (1954), and Debreu (1959), in which the commodity vector is defined to include dated claims to goods, possibly made contingent on random events. Prescott and I (Lucas and Prescott 1971) adopted this framework for the construction of a rational expectations model of investment in a competitive industry, taking a stochastically shifting demand curve (rather than prices) as given. And, in a paper that was to set the research agenda for the next decade, Kydland and Prescott (1982) utilized a version of the stochastic growth optimal growth model of Brock and Mirman (1972) as an operational model of a competitive economy undergoing recurrent business cycles, induced by shifts in the technology. This turned out to be a tremendously fruitful idea, whose potential is still being realized. But such a model without money is obviously not suited to the study of Hume's problem. Economists who believed that monetary forces were at the center of the business cycle needed to look elsewhere.

A second general equilibrium framework, due to Samuelson (1958), was also available and seemed better suited to the study of monetary questions. That paper introduced a deceptively simple example of an economy in which money with no direct use in either consumption or production nonetheless plays an essential role in economic life. I used this model (Lucas 1972*b*) in an attempt to show how monetary neutrality might be reconciled with the appearance of a short-term stimulus from a monetary expansion. The model is so simple and flexible that it can be used to illustrate many issues. I shall introduce a version of it here, along with enough notation to permit discussion of some interesting details.

In Samuelson's model, people live for two periods only, so that the ongoing economy is always populated by two age cohorts, one young and the other old. Here I assume a constant population, so that per capita and economywide magnitudes can be used interchangeably. At each year's end, the old die, the young become old, and a new young group arrives. It is important for my purposes (as it was for

Samuelson's) to assume that there is no family structure in this economy: no inheritances and no financial support by one cohort for another. Suppose that a young person in this economy can work and produce goods, whereas an old person likes to consume goods but has no ability to produce them. Denote a person's two objects of choice by the pair (c, n), where n is units of labor supplied when young and c is units of the good consumed when old. Assume that everyone's preferences over these two goods are given by $U(c) - n$. Assume a labor-only technology in which one unit of labor yields one unit of goods.

If the good were storable, everyone would produce in his youth and carry the production over for his own later consumption, solving the problem

$$\max_{n} [U(n) - n]. \tag{1}$$

Call the solution to this problem n^*. But I shall assume that the good cannot be stored, so that any individual acting purely on his own *cannot* produce for his own pleasure.[5] The best one acting alone can do is to enjoy leisure when young and never consume anything. Clearly society as a whole should be able to do *much* better than that, by somehow inducing the young to produce for the consumption of their *contemporary* old. Some institution is needed to achieve this.

A social security system may be one real-world instance of such an institution. (Or it may not: Everything hinges on the realism of the assumption of no family structure.) As Samuelson noted, a monetary system may be another such institution, for one can view the failure of the autarchic allocation as arising from the absence of the double coincidence of wants that barter exchange requires. Those who wish to consume goods, the old, have nothing to offer in return to those who are able to produce, the young. But suppose that there were some paper money in circulation, initially in the hands of the old. The old would offer this cash to the young in exchange for goods, establishing a market price of some kind. Would the young accept these tokens—intrinsically useless, in Wallace's (1980) terms—and hence keep the value of tokens in terms of goods at any level above zero? Maybe not: This possibility can certainly not be ruled out. If the young were willing to produce goods in exchange for fiat money, the reason would have to be that they hoped to be able to trade the money they received for goods in their own old age.

The interesting thing about Samuelson's example is that this second

[5] In fact, Cass and Yaari (1966) show that even if storage is possible, the autarchic allocation can be improved on since it ties up goods in inventory permanently and unnecessarily.

scenario cannot be ruled out. It is possible, though by no means necessary, that the money in this economy will circulate forever, being exchanged over and over again for goods. If this exchange takes place in a single competitive spot market and the price p is established, then a young person who begins with no money and works n units will acquire pn units of cash. If he spends it all on goods next period, this yields $pn/p = n$ units of consumption. Thus everyone solves the problem (1). If the money supply is constant and evenly distributed over the old in the amount m per person, the equilibrium price will also be constant, at the level $p = m/n^*$. Evidently, this equilibrium is quantity-theoretic in Hume's sense: if m is (somehow) increased, the equilibrium price level will be increased in the same proportion, and quantities of labor and production will not be affected at all.

When we consider monetary changes that differ from once-and-for-all changes in the money *stock*, however, the issue of neutrality becomes more complicated. To see this, suppose that we replace the assumption of a constant money supply with the assumption that the quantity of money grows at a constant percentage rate. We need to be explicit (another point in favor of Samuelson's model) about the way the new money gets into the system, and it matters how this is done. Assume, to begin with, that each young person receives an equal share of the newly created money, in between his youth and old age, and that the size of this addition to his cash is independent of the amount he has earned by working. Then if the money supply is m and is to be augmented by the lump-sum transfer $m(x - 1)$, each young person now solves

$$\max_{n} \left[U\left(\frac{pn + (x-1)m}{p'} \right) - n \right], \tag{2}$$

where p is the price at which he sells goods, today, and p' is the price at which he buys goods, tomorrow. The first-order condition for this problem is

$$U'\left(\frac{pn + (x-1)m}{p'} \right) \frac{p}{p'} = 1. \tag{3}$$

In order to work out a rational expectations equilibrium for this model, we exploit the observation that the only thing that changes over time in this situation is the money supply, which is simply multiplied by the known factor x in every period. It seems natural, then, to seek an equilibrium in which the price level is proportional to the money stock, $p = km$ for some constant k, and in which labor is constant at some value $\hat{n}$. In such an equilibrium, the constant k will evidently be $1/\hat{n}$. Tomorrow's price is then $p' = kmx = mx/\hat{n}$. Inserting all this information into the first-order condition (3), one obtains

$$U'(\hat{n}) = x. \tag{4}$$

In this circumstance, then, the price level will increase between periods at exactly the rate of growth of the money supply. The equilibrium level of employment $\hat{n}$, from (4), will be a decreasing function of the rate of money growth.[6]

The quantity-theoretic predictions we saw confirmed in figures 1 and 2 would also be confirmed in this hypothetical world. But note that this does *not* mean that the rate of money growth and the equal rate of price inflation are merely units changes, of no consequence to anyone. The faster money grows, the more important the overnight transfer is, relative to the cash accumulated through working. The monetary transfers dilute the return from working. Goods production declines as the inflation rate rises, and everyone is made worse off. This is a nonneutrality of money, a real effect of a money change (some would prefer to call it a real effect of the fiscal transfer that is used to bring the money change about), but this effect is obviously not the stimulating effect of a monetary expansion that Hume discusses. In this example, inflation does not "quicken the diligence of every individual." It is a kind of tax that deadens diligence by reducing its real return.

This inflation tax is an issue of the first importance, I think, and its effects are captured in a useful way by the theoretical example I have just worked through. But further study of the inflation tax is not going to bring us any closer to an understanding of the trade-off that Hume thought he observed, and that so many others have seen since. Let us then get the inflation tax out of the picture by assuming that the fiscal transfers through which the money supply expands are made in proportion to the balances one has earned through working. That is, if one works n units, one receives the transfer $pn(x - 1)$, not $m(x - 1)$, and thus has pnx to spend next period. In this situation, the first-order condition (4) becomes $U'(\hat{n}) = 1$, independent of x, and $\hat{n}$ is always at its efficient level n^*: there is no inflation tax. These proportional transfers are just an assumption of convenience, but one that will simplify the discussion of some hard questions.

Now how might this overlapping generations economy be modified so that a monetary expansion will act as a stimulus to production?

[6] Jörgen Weibull pointed out to me that one could obtain a version of this example in which equilibrium employment is an *increasing* function of the money shock x by assuming that only some of the young receive the entire transfer and by making the right assumptions about the curvature of the function U.

One might think that this could be achieved by replacing the assumption that the transfer variable x is constant with the assumption that it is drawn independently each period from some fixed probability distribution. Evidently, if the current-period realization is known to everyone, this will not change anything. What is perhaps less obvious, but equally true, is that even if the transfer realization is known directly only to the old, it will be revealed perfectly to the young by the equilibrium price that it induces. As in the constant money growth example we worked through above, prices are determined by m and x. What else is there in this context? If m is known and p is observed, as of course it must be in competitive trading, then one can infer the value of x.

In order to get an output effect from a monetary shock, then, it is not enough simply to introduce uncertainty. We need to imagine that the exchange of money for goods takes place in some manner other than in a centralized Walrasian market. In Lucas (1972*b*), I assumed that exchange occurs in two markets, each with a different number of goods suppliers. In this circumstance, a given price increase can signal a supplier that the money transfer x is large, in which case he wants to treat it like a units change and not respond; or it can mean that there are only a few suppliers in his market, in which case he wants to treat it like a real shift in his favor and respond by producing more. The best the individual can do, given his limited information, is to hedge. On average, then, labor supply and production are an increasing function $\varphi(x)$ of the monetary transfer. Equilibrium prices, $mx/\varphi(x)$, move in proportion to m, which is known to all traders, but increase less than proportionally with the transfer x. By next period, the transfer x is known, and prices complete their proportional increase, but not without a transition during which production is increased.

The resemblance of this scenario to the one I quoted from Hume in my Introduction seems clear. In an important sense the new scenario is an improvement since in place of the unexplained errors of judgment or ignorance that lie at the center of Hume's account, this one rests on an assumption that people lack complete information. But perhaps this only pushes the question back one step: *Why is it* that people cannot obtain that last bit of information that would enable them to diagnose price movements accurately? In reality, up-to-date information on the money supply does not seem all that hard to come by.

Let us step back from the specifics of this particular, information-based version of Hume's scenario and consider the possibilities more abstractly. Assume simply that old and young engage in some kind of trading game, to which the old bring the cash m obtained in the previous period's trading.[7] Either before or perhaps during the play of this game, the old receive a proportional transfer that totals x. Let each young person and each old person select a trading strategy. Notice that the strategy of a young person can depend on m, and the strategy of an old person can depend on m and x. On the basis of these choices, suppose that a Nash equilibrium is reached under which each young person supplies some amount of labor and ends up with some amount of cash. I shall restrict attention to symmetric equilibria, so that in equilibrium each young person ends up with mx dollars. Each young person also ends up supplying $f(m, x)$ units of labor, and this quantity is also the equilibrium consumption of each old person (the notation is chosen to emphasize that m and x are the only state variables in this model). Different specifications of the trading game will have different implications for this *outcome function f*.

Now assume that before the play of such a game begins, the money stock m is evenly distributed over the old; that everyone, young and old, knows what it is; and that everyone knows how transfers occur—the rules of this trading game. In these circumstances, changes in m must be neutral units changes, so that f is constant with respect to m and can be written $f(m, x) = \varphi(x)$ for some function φ. Given this function φ, the average price of goods is just the money stock divided by production, or $p = mx/\varphi(x)$. In competitive trading, φ is a constant function, so price is proportional to mx, where x is known; but in many other trading games, the function φ will vary with the value x. In this notation, rationalizing a trade-off of the type described by Hume translates into constructing a game that rationalizes an increasing function $\varphi(x)$.

One such game (though that equilibrium was not quite symmetric) was described in Lucas (1972*b*). There, the response in output was based on suppliers' imperfect information about the transfer x. But at this level of abstraction there are many other noncompetitive trading games that have outcomes with these same features. Some of them achieve this end by assuming that some nominal prices are set in advance, as in Fischer (1977), Phelps and Taylor (1977), Taylor (1979), or Svensson (1986). Others postulate games in which the transfer is only gradually revealed, as in Eden (1994), Lucas and Woodford (1994), or Williamson (1995). All these papers offer rationalizations of a short-run monetary nonneutrality in the sense of an increasing function $\varphi(x)$, though of course in quite different ways. In an important sense, then, Hume's paradox has been resolved: We have a wide variety of theories that reconcile long-run monetary neutrality with a short-run trade-off. They all (and any other game that

[7] This point of departure has long been advocated by Shubik (see, e.g., Shubik 1980).

fits into the formalism above) carry the implication that anticipated money changes will not stimulate production and that at least some unanticipated changes can do so.[8]

Does it matter which of these rationales is appealed to? The answer to this harder question must depend on what our purposes are. Any of these models leads to the distinction between anticipated and unanticipated changes in money, the distinction that seems to me the central lesson of the theoretical work of the 1970s. On the other hand, none of these models deduces the function φ from assumptions on technology and preferences alone. Of course φ depends on such factors, but it also depends on the specific assumptions one makes about the strategies available to the players, the timing of moves, the way in which information is revealed, and so on. Moreover, these specifics are all, for the sake of tractability, highly unrealistic and stylized: We cannot choose among them on the basis of descriptive realism. Consequently, we have no reason to believe that the function φ is invariant under changes in monetary policy—it is just a kind of Phillips curve, after all—and no reliable way to break it down into well-understood components.

Theories that emphasized the distinction between anticipated and unanticipated money shocks led to a variety of statistical tests. Sargent (1976) interpreted the prediction that anticipated money would have no real effects as the hypothesis that money would not "cause," in the sense of Granger (1969) and Sims (1972), changes in unemployment rates, and he found that this prediction was confirmed for U.S. time series. Barro (1977) used residuals from regressions of M1 on its own lagged values as measures of unanticipated money shocks and concluded that the unemployment rate responded to these shocks but did not respond to current and lagged M1.[9] The signal-processing feature of the model of Lucas (1972*b*) implied that the magnitude of a money multiplier should decline as the variance of money changes increased. This prediction was confirmed in the cross-country comparisons reported in Lucas (1973) and Alberro (1981) and by the much more extensive results reported in Kormendi and Meguire (1984).

In the models in Lucas (1972*b*, 1973), trade takes place in competitive markets, though these markets are incomplete; so any real effects of monetary policy need to work through movements in prices. The tests described in the last paragraph do not use data on prices and so do not test this prediction. Other econometric work that did re-

[8] Of course, this conclusion requires the usual caveat about the inflation tax.

[9] Whether this work in fact tests implications of the model in Lucas (1972*b*) is questioned in King (1981).

quire money shocks to be transmitted through price movements was much less favorable. Estimates in Sargent (1976) and in Leiderman (1979) indicated that only small fractions of output variability can be accounted for by unexpected price movements. Though the evidence seems to show that monetary surprises have real effects, they do not seem to be transmitted through price surprises, as in Lucas (1972*b*).[10]

V. Conclusions

The main finding that emerged from the research of the 1970s is that anticipated changes and unanticipated changes in money growth have very different effects. Anticipated monetary expansions have inflation tax effects and induce an inflation premium on nominal interest rates, but they are not associated with the kind of stimulus to employment and production that Hume described. Unanticipated monetary expansions, on the other hand, can stimulate production as, symmetrically, unanticipated contractions can induce depression. The importance of this distinction between anticipated and unanticipated monetary changes is an implication of every one of the many different models, all using rational expectations, that were developed during the 1970s to account for short-term trade-offs. This distinction is consistent with the long-run evidence displayed in figures 1 and 2, with the year-to-year changes displayed in figure 3, with Friedman and Schwartz's account of depressions in the United States, and with Sargent's account of the ending of the European hyperinflations.

The discovery of the central role of the distinction between anticipated and unanticipated money shocks resulted from the attempts, on the part of many researchers, to formulate mathematically explicit models that were capable of addressing the issues raised by Hume. But I think it is clear that none of the specific models that captured this distinction in the 1970s can now be viewed as a satisfactory theory of business cycles. Perhaps in part as a response to the difficulties with the monetary-based business cycle models of the 1970s, much recent research has followed the lead of Kydland and Prescott (1982) and emphasized the effects of purely real forces on employment and production.[11] This research has shown how general equilibrium rea-

[10] Wallace (1992) develops a variation of the Lucas (1972*b*) model in which real shocks need not be perfectly negatively correlated across markets (so that real shocks can be positive in the aggregate). In this more general model, money shocks can induce output movements in the same direction (but not perfectly correlated) and the inflation-output correlation can have either sign. The evidence in Sargent (1976) and Leiderman (1979) is not decisive against such a variation.

[11] The approach of directly testing the implications of Euler equations, initiated in Hall (1978) and advanced by methods described in Hansen (1982), has also been productively pursued.

soning can add discipline to the study of an economy's distributed lag response to shocks, as well as to the study of the nature of the shocks themselves. More recently, many have tried to reintroduce monetary features into these models, and I expect much future work in this direction.

But who can say how the macroeconomic theory of the future will develop, any more than anyone in 1960 could have foreseen the developments I have described in this lecture? All one can be sure of is that progress will result from the continued effort to formulate explicit theories that fit the facts, and that the best and most practical macroeconomics will make use of developments in basic economic theory.

References

Alberro, José. "The Lucas Hypothesis on the Phillips Curve: Further International Evidence." *J. Monetary Econ.* 7 (March 1981): 239–50.

Arrow, Kenneth J. "An Extension of the Basic Theorems of Classical Welfare Economics." In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability,* edited by Jerzy Neyman. Berkeley: Univ. California Press, 1951.

Barro, Robert J. "Unanticipated Money Growth and Unemployment in the United States." *A.E.R.* 67 (March 1977): 101–15.

Baumol, William J. "The Transactions Demand for Cash: An Inventory Theoretic Approach." *Q.J.E.* 66 (November 1952): 545–56.

Becker, Gary S. "Investment in Human Capital: A Theoretical Analysis." *J.P.E.* 70, no. 5, pt. 2 (October 1962): 9–49.

Bellman, Richard E. *Dynamic Programming.* Princeton, N.J.: Princeton Univ. Press, 1957.

Ben-Porath, Yoram. "The Production of Human Capital and the Life Cycle of Earnings." *J.P.E.* 75, no. 4, pt. 1 (August 1967): 352–65.

Brock, William A., and Mirman, Leonard J. "Optimal Economic Growth and Uncertainty: The Discounted Case." *J. Econ. Theory* 4 (June 1972): 479–513.

Brunner, Karl, and Meltzer, Allan H. "Predicting Velocity: Implications for Theory and Policy." *J. Finance* 18 (May 1963): 319–54.

Cass, David. "Optimum Growth in an Aggregative Model of Capital Accumulation." *Rev. Econ. Studies* 32 (July 1965): 233–40.

Cass, David, and Yaari, Menahem E. "A Re-examination of the Pure Consumption Loans Model." *J.P.E.* 74 (August 1966): 353–67.

Debreu, Gerard. *Theory of Value: An Axiomatic Analysis of Economic Equilibrium.* New Haven, Conn.: Yale Univ. Press, 1959.

Eden, Benjamin. "The Adjustment of Prices to Monetary Shocks When Trade Is Uncertain and Sequential." *J.P.E.* 102 (June 1994): 493–509.

Eisner, Robert, and Strotz, Robert H. "Determinants of Business Investment." In *Impacts of Monetary Policy,* by the Commission on Money and Credit. Englewood Cliffs, N.J.: Prentice-Hall, 1963.

Fama, Eugene F. "The Behavior of Stock-Market Prices." *J. Bus.* 38 (January 1965): 34–105.

Fischer, Stanley. "Long-Term Contracts, Rational Expectations, and the Optimal Money Supply Rule." *J.P.E.* 85 (February 1977): 191–205.

Friedman, Milton. *A Theory of the Consumption Function.* Princeton, N.J.: Princeton Univ. Press (for NBER), 1957.

———. "The Role of Monetary Policy." *A.E.R.* 58 (March 1968): 1–17.

Friedman, Milton, and Schwartz, Anna J. *A Monetary History of the United States, 1867–1960.* Princeton, N.J.: Princeton Univ. Press (for NBER), 1963.

Granger, Clive W. J. "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods." *Econometrica* 37 (July 1969): 424–38.

Hall, Robert E. "Stochastic Implications of the Life Cycle–Permanent Income Hypothesis: Theory and Evidence." *J.P.E.* 86 (December 1978): 971–88.

Hansen, Lars Peter. "Large Sample Properties of Generalized Method of Moments Estimators." *Econometrica* 50 (July 1982): 1029–54.

Hayek, Friedrich A. von. *Monetary Theory and the Trade Cycle.* London: Jonathan Cape, 1933.

Hicks, John R. *Value and Capital.* Oxford: Clarendon, 1939.

Hume, David. *Writings on Economics,* edited by Eugene Rotwein. Madison: Univ. Wisconsin Press, 1970.

Jorgenson, Dale W. "Capital Theory and Investment Behavior." *A.E.R. Papers and Proc.* 53 (May 1963): 247–59.

Keynes, John Maynard. *A Treatise on Money.* London: Macmillan, 1930.

———. *The General Theory of Employment, Interest, and Money.* London: Macmillan, 1936.

King, Robert G. "Monetary Information and Monetary Neutrality." *J. Monetary Econ.* 7 (March 1981): 195–206.

Kormendi, Roger C., and Meguire, Philip G. "Cross-Regime Evidence of Macroeconomic Rationality." *J.P.E.* 92 (October 1984): 875–908.

Kydland, Finn E., and Prescott, Edward C. "Time to Build and Aggregate Fluctuations." *Econometrica* 50 (November 1982): 1345–70.

Leiderman, Leonardo. "Expectations and Output-Inflation Tradeoffs in a Fixed-Exchange-Rate Economy." *J.P.E.* 87 (December 1979): 1285–1306.

Lucas, Robert E., Jr. "Econometric Testing of the Natural Rate Hypothesis." In *The Econometrics of Price Determination: Conference, October 30–31, 1970, Washington, D.C.,* edited by Otto Eckstein. Washington: Board of Governors, Fed. Reserve System, 1972. (*a*)

———. "Expectations and the Neutrality of Money." *J. Econ. Theory* 4 (April 1972): 103–24. (*b*)

———. "Some International Evidence on Output-Inflation Tradeoffs." *A.E.R.* 63 (June 1973): 326–34.

———. "Econometric Policy Evaluation: A Critique." *Carnegie-Rochester Conf. Ser. Public Policy,* no. 1 (1976), pp. 19–46.

Lucas, Robert E., Jr., and Prescott, Edward C. "Investment under Uncertainty." *Econometrica* 39 (September 1971): 659–81.

Lucas, Robert E., Jr., and Rapping, Leonard A. "Real Wages, Employment, and Inflation." *J.P.E.* 77 (September/October 1969): 721–54.

Lucas, Robert E., Jr., and Woodford, Michael. "Real Effects of Monetary Shocks in an Economy with Sequential Purchases." Working paper. Chicago: Univ. Chicago, 1994.

McCandless, George T., Jr., and Weber, Warren E. "Some Monetary Facts." *Fed. Reserve Bank Minneapolis Q. Rev.* 19 (Summer 1995): 2–11.

McKenzie, Lionel W. "On Equilibrium in Graham's Model of World Trade and Other Competitive Systems." *Econometrica* 22 (April 1954): 147–61.

Meltzer, Allan H. "The Demand for Money: The Evidence from the Time Series." *J.P.E.* 71 (June 1963): 219–46.

Modigliani, Franco, and Brumberg, Richard. "Utility Analysis and the Consumption Function: An Interpretation of Cross-Section Data." In *Post-Keynesian Economics*, edited by Kenneth K. Kurihara. New Brunswick, N.J.: Rutgers Univ. Press, 1954.

Muth, John F. "Rational Expectations and the Theory of Price Movements." *Econometrica* 29 (July 1961): 315–35.

Patinkin, Don. *Money, Interest, and Prices: An Integration of Monetary and Value Theory*. 2d ed. New York: Harper and Row, 1965.

Phelps, Edmund S. "Money-Wage Dynamics and Labor-Market Equilibrium." *J.P.E.* 76, no. 4, pt. 2 (July/August 1968): 678–711.

Phelps, Edmund S., et al. *Microeconomic Foundations of Employment and Inflation Theory*. New York: Norton, 1970.

Phelps, Edmund S., and Taylor, John B. "Stabilizing Powers of Monetary Policy under Rational Expectations." *J.P.E.* 85 (February 1977): 163–90.

Phillips, A. W. "The Relation between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom, 1861–1957." *Economica* 25 (November 1958): 283–99.

Samuelson, Paul A. "An Exact Consumption-Loan Model of Interest with or without the Social Contrivance of Money." *J.P.E.* 66 (December 1958): 467–82.

Sargent, Thomas J. "A Note on the 'Accelerationist' Controversy." *J. Money, Credit and Banking* 3 (August 1971): 721–25.

———. "A Classical Macroeconometric Model for the United States." *J.P.E.* 84 (April 1976): 207–37.

———. "The Ends of Four Big Inflations." In *Rational Expectations and Inflation*. New York: Harper and Row, 1986.

Shubik, Martin. "The Capital Stock Modified Competitive Equilibrium." In *Models of Monetary Economies*, edited by John H. Kareken and Neil Wallace. Minneapolis: Fed. Reserve Bank, 1980.

Sims, Christopher A. "Money, Income, and Causality." *A.E.R.* 62 (September 1972): 540–52.

Stockman, Alan C. *Introduction to Economics*. Fort Worth, Texas: Dryden, 1996.

Svensson, Lars E. O. "Sticky Goods Prices, Flexible Asset Prices, Monopolistic Competition, and Monetary Policy." *Rev. Econ. Studies* 53 (July 1986): 385–405.

Taylor, John B. "Estimation and Control of a Macroeconomic Model with Rational Expectations." *Econometrica* 47 (September 1979): 1267–86.

Tinbergen, Jan. *Business Cycles in the United States of America, 1919–1932*. Geneva: League of Nations, 1939.

Tobin, James. "The Interest-Elasticity of Transactions Demand for Cash." *Rev. Econ. and Statis.* 38 (August 1956): 241–47.

Uzawa, Hirofumi. "Optimal Growth in a Two-Sector Model of Capital Accumulation." *Rev. Econ. Studies* 31 (January 1964): 1–24.

Wallace, Neil. "The Overlapping Generations Model of Fiat Money." In *Models of Monetary Economies*, edited by John H. Kareken and Neil Wallace. Minneapolis: Fed. Reserve Bank, 1980.

———. "Lucas's Signal-Extraction Model: A Finite State Exposition with Aggregate Real Shocks." *J. Monetary Econ.* 30 (December 1992): 433–47.

Wicksell, Knut. *Interest and Prices: A Study of the Causes Regulating the Value of Money*. 1898. Translated by R. F. Kahn. London: Macmillan (for Royal Econ. Soc.), 1936.

Williamson, Stephen D. "Optimal Monetary Policy in an Economy with Sequential Service." Working paper. Iowa City: Univ. Iowa, 1995.