How Powerful Is Monetary Policy in the Long Run?

MARCO A. ESPINOSA-VEGA

The author is a senior economist in the macropolicy section of the Atlanta Fed's research department. He thanks Frank King for comments on an earlier version of this article and Steve Russell for detailed and thoughtful editorial suggestions.

RESS REPORTS ABOUT THE STATE OF THE ECONOMY OFTEN GIVE READERS THE IMPRESSION THAT MONETARY POLICY AND THE PEOPLE WHO DIRECT IT ARE QUITE POWERFUL. FOR EXAMPLE, AN ARTICLE IN THE Washington Post in March 1997 asserts that "second to the president, Alan Greenspan is arguably the nation's most powerful person. As chairman of the Fed, he guides U.S. Monetary policy, adjusting short-run interest rates."

Many prominent academic economists seem to agree that monetary policy is quite powerful. In reviewing the monetary policy experience of the 1970s, Nobel Laureate James Tobin wrote, "In one respect demand-management policies worked as intended in the 1970s. . . . the decade is distinguished by its three recessions, all deliberately induced by policy. Likewise the expansionary policies adopted to reverse the first two recessions, beginning in 1971 and 1975 respectively, promoted recoveries and in 1977 expansion . . . The major turns in direction conformed to the desires and intentions of the managers of aggregate demand" (1980, 20–21).

Monetary policymakers themselves often describe their role as powerful. Consider, for example, Federal Reserve Chairman Alan Greenspan's testimony before Congress in July of last year. Attempting to explain the influence of monetary policy on the current state of the economy, he stated that "the preemptive actions of the Federal Reserve in 1994 contained a potentially destabilizing surge in demand, short-circuiting a boom-bust business cycle in the making" (1997). Without attempting to explain the full meaning of Greenspan's statement here, it is clear from his language that he believes the Federal Reserve System is powerful enough to have a profound influence on the course of economic activity.

Both Greenspan's statement and Tobin's comments focus on the short-run effects of monetary policy. One might suspect that if Greenspan really is the second most powerful person in the United States then the policy tools he controls must have some long-run influence on the U.S. economy. Ironically, however, although many academic economists and most Federal Reserve policy-makers believe that monetary policy is quite powerful in the short run, they also believe that it is virtually powerless in the long run.

Although opinion on this subject is far from uniform, most economists seem to believe that monetary policy can affect the level of real (inflation-adjusted) economic activity—that is, economic variables such as real interest rates, real gross domestic product (GDP) and the unemployment rate—over periods of one or two years. For example, the Fed can create economic recessions or strengthen cyclical expansions. It can do so, according to the conventional view, by increasing the growth rate of the money supply if it wants the economy to grow faster and reducing it if it wants the pace of economic activity to slow. However, increases in the money supply growth rate eventually cause the inflation rate to rise, and decreases in the money growth rate have the opposite effect. When policymakers discuss the short-

run effects of monetary policy they usually describe some version of this trade-off between higher inflation, which almost everyone considers undesirable, and desirable changes in other economic variables: higher inflation vs. lower interest rates, lower unemployment, or faster growth in real GDP.

As indicated, however, most economists believe that the long-run effects of changes in monetary policy are very different from their short-run effects. Federal Reserve Governor Meyer has clearly stated the nature of this difference in beliefs, commenting that "there is, to be sure, no trade-off and hence no inconsistency between full employment and price stability in the long run" (1997, 19).

A pair of simple diagrams illustrates the conventional views about the short- and long-run effects of monetary policy. Chart 1 depicts a negatively sloping curve that describes a short-run trade-off between inflation and unemployment. In contrast, Chart 2, which displays a vertical line at the level of full employment, illustrates a scenario in which there is no trade-off between the level of unemployment and the rate of inflation. A low rate of inflation (price stability) is compatible with full employment, but so is a high rate of inflation. If Chart 2 accurately describes the long-run relationship between unemployment and inflation, then changes in monetary policy that lead to changes in the inflation rate have no effect on the long-run levels of unemployment or real output. In the jargon of economists, this diagram describes a situation in which money is superneutral in the long run.²

Although the view that monetary policy has real effects in the short run but is superneutral in the long run is widely accepted by academic economists, business economists, and economic policymakers, these groups are not in complete agreement about the ultimate real effects of monetary policy. One source of differences involves the magnitude of the short-run effects. Business economists and policymakers tend to believe that the short-run effects of monetary policy are very large, but most of their academic counterparts see these effects as rather tame and inconsequential. A related difference involves the questions of whether any short-run power that the Fed may have can survive repeated use. Most nonacademics seem to believe that the Fed can use its policy tools as often as it wishes

without fear that they will lose their short-run effectiveness. On the other hand, most academic economists believe that repeated, systematic efforts to use the Fed's power to affect real economic activity will grow less and less effective over time.

This article reviews the development of the consensus view that monetary policy can have short-run effects but that it is long-run superneutral. The discussion

emphasizes the fact that the basis for this view is the assumption that the only way monetary policy can affect real economic activity is via "money illusion"—that is, by creating changes in the price level that are misunderstood by households and firms and cause them to make bad economic decisions. If monetary policy can affect real economic activity by means other than money illusion then it

Most economists seem to believe that monetary policy can affect the level of real (inflation-adjusted) economic activity over periods of one or two years.

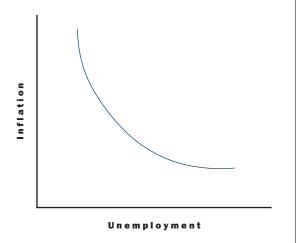
may be possible for money to be nonsuperneutral in the long run.

This article hopes to challenge economists and policymakers to devote more attention to investigating alternative explanations for the real effects of monetary policy—explanations that may imply that money is not long-run superneutral. In order to develop these alternative explanations it is necessary to make very explicit assumptions about the role of money in an economy, how it interacts with real variables and how economic decisionmakers react to the changes in its supply. Different assumptions will turn out to have very important implications for both the nature and the magnitude of the results of policy changes. This point is illustrated in the review of the small but growing branch of the academic literature on the real effects of monetary policy literature that studies models in which money may not be long-run superneutral. In these models the ultimate source of the real effects of monetary policy is the credit markets. By linking monetary policy with the supply of credit these models can analyze an alternative

Linton Weeks and John Berry, "The Shy Wizard of Money: Fed's Enigmatic Greenspan Moves Easily in His Own World," Washington Post, March 24, 1997, sec. A.

^{2.} Money is said to exhibit long-run neutrality if permanent changes in the level of the supply of money have no long-run effects on real interest rates or the growth rate of real output. In this case, the levels to which prices and other nominal variables will increase are postulated to vary one for one with changes in the level of the money supply. Similarly, an economy is said to display long-run superneutrality if permanent changes in the rate of growth of the money supply have no long-run effects on either real interest rates or the rate of output growth, and the rates of inflation and other nominal variables are postulated to vary one for one with changes in the rate of growth of the money supply.

CHART 1 A Short-Run InflationUnemployment Trade-Off



mechanism for evaluating the long-run effects of monetary policy that does not rely on money surprises.

Another important message of this article is that the very idea that monetary policy is powerful in the short run but powerless in the long run may be internally inconsistent.³ If monetary policy is indeed as powerful as many informed people seem to believe, then theories of its real effects that rely on money illusion may have to be replaced by theories in which money is not superneutral in the long run.

The Precursors

his section briefly reviews the evolution of two prominent views on the neutrality of money: the Keynesian view and the monetarist view. The discussion begins with a look back at the classical theory that preceded Keynesianism and monetarism. It concludes by describing the clash between the Keynesians and the monetarists and the resulting "unilateral synthesis" of the 1970s.

The Early Quantity Theory. Classical macroeconomic theory, which developed during the late nineteenth and early twentieth centuries, was characterized by its focus on economic fundamentals ("real" economic conditions) such as individuals' propensity to save, the state of technology, and so on. In the classical view monetary policy played no long-run role in determining real economic activity. In particular, it had no long-run effect on the level of real interest rates. Classical theorists acknowledged that monetary policy might have a minor influence over economic activity (particularly interest rates) in the short run. In the long run, however, they viewed money as having a direct influence only on prices.

This early view of the influence of money on prices came to be known as the quantity theory of money. Like

many economic concepts, the quantity theory has a rich history of reinterpretations. One of the earliest statements of the theory in its modern form was presented by Fisher (1926). According to Fisher, an economy's general price level is a function of the quantity of money in circulation, the economy's efficiency, or velocity, of circulation (the average number of times a year money is exchanged for goods), and the volume of trade (the quantity of goods purchased with money). Notationally, Fisher expresses the equation of exchange as:

$$M \times V = P \times T$$

where M is the supply of money, V is the velocity of money, P is the general price level, and T is the total value of transactions or trade. Fisher held that in the long run there was a "natural" level of real economic activity determined by economic fundamentals that could not be affected by increases in the amount of money in the economy. In his words, "An inflation of the currency cannot increase the product of farms and factories . . . The stream of business depends on natural resources and natural conditions, not on the quantity of money" (1926, 155). This hypothesis that there was a natural long-run level of real economic activity, together with the assumption that the only role of money is to serve as a unit of account, formed the basis of Fisher's quantity theory analysis that prices varied proportionately to changes in the quantity of money. According to this equation, if velocity of money and the value of transactions were fairly stable—at least in the long run—as the economy approached its natural level, then changes in the quantity of money would be met with proportional changes in the price level.

Fisher conceded that monetary policy might have some temporary effects on real economic activity, commenting that "the 'quantity theory' will not hold true strictly and absolutely during transition periods" (1926, 161). In his mind, however, these effects were mainly due to temporary changes in velocity. If velocity was fairly stable in the long run, though, as he assumed, then it had to be the case that prices varied proportionately with the supply of money.

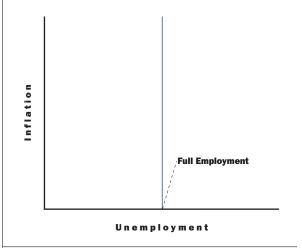
As this description indicates, the basic current consensus on the short- and long-run effects of monetary policy can be traced to the early quantity theorists. According to their view, as represented by Fisher, monetary policy could have temporary real effects but it would be superneutral in the long run. However, even the theory's adherents understood that the theory needed further refinement. This task was undertaken a few years later by Milton Friedman (see below). By the time of the Great Depression, moreover, classical theory had lost much of its popularity and a new, completely different macroeconomic theory was appearing.

The Keynesians and Money: The First Time Around.

The first nonclassical macroeconomic theory was the creation of John Maynard Keynes and is laid out in his General Theory (1936). One of Keynes's principal goals was to identify the causes of the persistently high rates of unemployment that were afflicting virtually the entire world during the Great Depression. He also sought to identify government policies that could help reduce these high levels of unemployment. Although Keynes's theory discussed the long-term implications of government policies, his focus was on the short run. And although monetary factors played a role in determining real economic activity in his theory (unlike in classical theory), Keynes's analysis emphasized fiscal policy. Keynes believed fiscal policy was the most powerful tool a government could use to lift the economy out of a recession or depression. In fact, his theory predicted that under certain conditions increases in the money supply would be unable to drive interest rates down low enough to stimulate economic activity by generating additional demand for credit. This situation was known as the liquidity trap. In liquidity trap situations money was superneutral even in the short run.

Nonneutrality of Money in the Long Run: The Chicago School. For many years after the Depression and the world war that followed it the question of the long-run implications of government policies received very little attention. One of the early assessments of the long-run effects of monetary policy came from, of all places, the University of Chicago. The university's department of economics—which was and remains perhaps the world's most influential collection of academic economists—has always been associated with the economic principles of the classical economists. The Chicago economics department was instrumental in the development of monetarism, which is usually considered to be a direct descendant of classical macroeconomic theory. In 1951, however, Lloyd Metzler, a prominent member of the economics faculty at Chicago, published a paper describing the long-run implications of central bank open market operations in which he asserted that under some circumstances monetary factors could interact with real variables in such a way as to help determine the level of real economic activity in both the short run and the long run. Metzler wrote that "by purchasing or selling securi-

CHART 2 An Inflation-Unemployment Relationship with No Trade-Off



ties, the banking authorities can alter not only the temporary interest rate which prevails while the open-market transaction is taking place but also the rate at which the system will return to equilibrium after the bank's transactions in securities have ceased" (1951, 107). He continued, "By purchasing securities, the central bank can... [cause] the system to attain a new equilibrium at a permanently lower interest rate and a permanently higher rate of capital accumulation" (112).

It is important to note that Metzler's conclusion that monetary policy—induced changes in the government's portfolio of liabilities could potentially have long-run real effects does not rely on the monetary authority's ability to produce inflation surprises or on workers' or firms' inability to correctly appraise conditions in the labor market. His analysis is therefore very different from most modern analyses, which view monetary surprises and their impact on naive participants in labor markets or short-run market frictions as the main channel by which monetary policy can affect real economic variables.

Although Metzler's analysis is less than fully rigorous by modern standards, it is worth recalling because it represents one of the first careful descriptions of a mechanism through which monetary policy can have long-run real effects. As noted above, Metzler's conclusions ran counter to the classical tradition of the Chicago school.

^{3.} If monetary policy has real effects only because of money illusion then it is likely these effects will be very limited in scale. On the other hand, if monetary policy derives its real effects from other sources, then its short-run effects may be relatively large. Thus, limited short-run real effects may go hand in hand with long-run superneutrality while deviations from long-run superneutrality may produce powerful short-run effects.

^{4.} In particular, as stated, this description of the quantity theory is really more of an accounting identity than a theory that qualitatively relates money to relevant macroeconomic variables. An accounting identity does not specify what is a given in the analysis and how different variables will change as a result of alternative policy changes. A theory or model, on the other hand, is specific about what is assumed to be exogenous to the model as well as what is determined within it and how different variables react to exogenous changes. A number of complimentary assumptions were really necessary for this equation to spell the list of properties Fisher attached to the quantity theory.

Perhaps for this reason Metzler's ideas failed to stimulate a research program at Chicago (or elsewhere). Instead, Chicago's monetarists pressed ahead with refinements of the quantity theory of money.

The Monetarists and the Modern Quantity Theory. During the 1950s the monetarists attempted to recover the popularity that classical theory had lost as a result of the Great Depression and the development of Keynesian macroeconomics. The most prominent monetarist was (and remains) Milton Friedman, an economist at the

In the classical view monetary policy played no longrun role in determining real economic activity. University of Chicago. One of Friedman's first major contributions to monetarism was a refinement of Fisher's quantity theory.

As Friedman pointed out, both the size of the money supply and the general level of prices can be considered public knowledge. However, for a theory to be able to make predictions regarding the effects of changes in monetary policy or

changes in the price level, it is necessary to establish some assumptions about what differentiates the behavior of money supply from the behavior of money demand. In Friedman's words, "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 with some specifications about the conditions of the supply of money . . . " (1956, 4).

Friedman's version of the quantity theory is based on the postulate that there is a stable demand for real money balances—that is, for purchasing power in monetary form. He assumes that in the long run the level of money demand depends on economic fundamentals such as real income, the interest rate, and the nature of the technology for conducting transactions. Under this assumption, changes in the nominal supply of money engineered by the Fed have no long-run impact on the real demand for money and consequently lead inevitably and exclusively to changes in the price level. This observation is true both for one-time changes in the money supply and for changes in the rate at which the money supply is growing, which would result in changes in the inflation rate but not in the levels or growth rates of real variables. Thus, one implication of Friedman's restatement of the quantity theory of money is that changes in monetary policy would have no real effects in the long run—that is, money would be long-run superneutral.

Money in Keynesian Analysis: The Second Time

Around. As the discussion has shown, early Keynesians focused their attention on fiscal policy. They believed that under normal circumstances changes in the general price level would be both infrequent and relatively inconsequential. As a result, for many years after the Second World War monetarists enjoyed a virtual monopoly over monetary analysis. This situation changed in the mid-1960s, when Keynesians developed a strong interest in the role of monetary policy.

Keynes himself rejected the quantity theory approach to determining the price level. For Keynes, the magnitude of the money supply in the economy was only one of a number of factors affecting the general level of prices. Another important factor was the level of employment. In Keynes's view it was impossible to determine the ultimate impact of a change in the quantity of money on the price level without considering the economy's overall level of employment. More specifically, Keynes believed that "an increase in the quantity of money will have no effect whatever on prices, so long as there is any unemployment" (1964, 295). Since Keynes saw persistent unemployment as the central problem facing industrialized economies, he did not think it would be unusual for economies to go for extended periods of time without observing significant changes in the price level. During the 1950s the general price level was indeed fairly stable. This circumstance lent credence to the Keynesian view that focusing on fiscal policies that might help solve chronic unemployment problems was likely to be more fruitful than devoting a lot of energy to analysis of price level determination.

As the postwar era wore on, inflation began to pick up in both the United States and Western Europe. This development stimulated interest in analyzing the causes of and cures for inflation. In 1958 British economist A.W. Phillips published an empirical analysis of historical data for the U.K. labor market. He hoped to find empirical support for the Keynesian hypothesis that the rate of wage inflation depended on the tightness of the labor market. Phillips found that from 1861 to 1957 the growth rate of nominal wages was negatively related to the rate of unemployment. This "Phillips curve" seemed to link the real side of the economy (the rate of unemployment) to the nominal side (nominal wages). And since wages are the biggest single component of firms' costs, most economists were willing to assume that persistent increases in wage rates would eventually force firms to begin increasing their prices, producing economywide price inflation.⁵

Although Phillips's findings were empirical in nature, they have had a profound and lasting effect on the development of economic theories about the relationship between inflation and real economic variables. As the discussion has shown, Keynesian theory holds that it is possible to use fiscal or monetary policy to increase or decrease the level of aggregate demand and through it

the level of employment. The Phillips curve created a link between the level of aggregate demand and the rate of inflation. As a result economic policymakers began to think of demand management policies as involving a trade-off between the unemployment rate (and, more generally, the level of real economic activity) and the inflation rate. And if the Phillips curve was stable over time then this trade-off would exist in both the short run and the long run.

Long-Run Nonsuperneutrality of Money: The **Keynesian School.** The first attempt to formalize the Keynesian view about the long-run real effects of monetary policy was presented by James Tobin. Unlike the classical economists (but like Metzler), Tobin saw real economic activity in general, and real interest rates in particular, as being determined jointly by economic fundamentals and by monetary policy—even in the long run. In Tobin's words, "Keynes gave reasons why in the short run monetary factors and portfolio decisions modify, and in some circumstances dominate, the determination of the interest rate and the process of capital accumulation. I have tried to show here that a similar proposition is true for the long run. The equilibrium interest rate and degree of capital intensity are in general affected by monetary supplies and portfolio behavior, as well as by technology and thrift" (1965, 684).

Tobin's analysis resembled Metzler's in abstracting from labor markets and concentrating on portfolio adjustments as the channel by which monetary policy could have long-run real effects. According to Tobin's theory, both money and physical capital were elements of an individual's portfolio of savings. For a given real rate of return on capital, an increase in the rate of inflation would make money less attractive and capital more attractive, inducing individuals to reduce their holdings of money in favor of holdings of physical capital. As a consequence, one would observe additional accumulation of capital, a higher capital stock, and a higher output level in the long run.

Long-Run Superneutrality of Money: The Monetarist School. What was the monetarist reaction to Keynesians' claim about the long-run effects of monetary policy? Monetarists did not address Tobin's arguments directly. Instead, they attempted to provide a theoretical underpinning for empirical work of the type conducted by Phillips (1958) that analyzed the relationship between nominal and real variables. Once the theoretical framework was in place, the monetarists used it to explain why monetary policy–induced changes in real economic activity would be short-lived.

While Phillips's statistical evidence involved nominal wages, standard economic theory assumes that house-

holds and firms base their employment decisions on real (inflation-adjusted) variables such as real wages, real interest rates, real profits, and so forth. Thus, additional assumptions were needed to reconcile standard economic theory with Phillips's findings. Ironically, the point of departure for this reconciliation was Keynes's observation that "every trade union will put up some resistance to a cut on money wages, however small, but no trade union would dream of striking on every occasion of a rise in the cost of living" (1964, 14–15). Friedman (1968) and Phelps

(1967) used Keynes's observation in an attempt to extract some economic content from the statistical relationship discovered by Phillips. Their explanation for the behavior Keynes described was based on two assumptions—one about the nature of monetary policy, and the other about economic decisionmaker's responses to the effects of monetary policy. The first assump-

One implication of
Friedman's restatement
of the quantity theory of
money is that changes in
monetary policy would have
no real effects in the long
run—that is, money would
be long-run superneutral.

tion is that increases in the money supply often cause "monetary surprises"—unexpected increases in the rate of inflation. The second assumption was that economic decisionmakers' reaction to monetary surprises often involves temporary money illusion, which is a temporary failure to recognize that there has been an increase in the price level. The basic idea here is that although monetary surprises increase the prices of all goods and services, economic decisionmakers usually notice the effects of these increases on particular prices in which they have a special interest—their wages or the prices of the goods they produce—well before they notice their effects on the overall price level. Until they discover that the overall price level has increased, they mistakenly believe that the increases in the money (nominal) prices of the goods they care about represent increases in the real prices (relative prices) of those goods. This mistaken belief can lead households or firms to make decisions about saving, consumption, work effort, investment, production, and so forth that are quite different from the decisions they would have made otherwise. As a result, by creating monetary surprises monetary policy can influence the level of real economic activity.

Here is a hypothetical sequence of events that illustrates how temporary money illusion can empower

^{5.} For this explanation to make sense some additional assumptions are required. See Espinosa and Russell (1997) for an explanation of these assumptions.

monetary policy: Suppose the economy starts out in its long-run equilibrium at its normal inflation rate and its "natural" real rate of interest. Suppose further that the monetary authority begins to increase the money supply at a faster pace than previously. The most immediate consequence of this move will be a drop in nominal interest rates. Friedman explains, "Let the Fed set out to keep interest rates down. How will it try to do so? By buying

Tobin saw real economic activity as being determined jointly by economic fundamentals and by monetary policy—even in the long run.

securities. This raises their prices and lowers their yields . . . In the process, it also increases . . . the total quantity of money. The initial impact of increasing the quantity of money at a faster rate than it has been increasing is to make interest rates lower for a time than they would otherwise have been" (1968, 5–6).

The next step in Friedman's chain of causation is that lower

interest rates will stimulate spending, and this increase in spending will have a multiplier effect on the overall level of economic activity. Friedman writes, "The more rapid rate of monetary growth will stimulate spending . . . one man's spending is another man's income" (1968, 5-6). From this point, Friedman's analysis can be illustrated using the aggregate demand and aggregate supply (AD and AS) diagram that appears in many textbooks in introductory macroeconomics. The economy starts out in a long-run equilibrium at the intersection of the AD and AS curves. The intersection point represents the long-run equilibrium levels of real output and the price level. In the AD-AS framework, a change in a variable like the market interest rate leads to changes in the market environment that determined the location of the AD and AS curves and consequently produces a shift in at least one of these curves. In this case, the increase in spending that results from the decline in interest rates (which was caused by the increase in the money supply growth rate) is represented by a rightward shift in the AD curve. This increase in aggregate demand produces an increase in output and prices along the original aggregate supply curve.

According to Friedman, this change in the equilibrium will be strictly a short-run phenomenon. As soon as households and firms realize that lower interest rates and faster-rising wages and product prices are also associated with a more rapid rate of increase in the overall price level—as soon, that is, as they realize that real wages and prices have not changed—the house-

holds will reduce their supply of labor and the firms will cut back their production. On the diagram, this behavior is represented by a leftward shift in the aggregate supply schedule that exactly offsets the effects of the increase in aggregate demand. In the end, the economy will return to the original long-run natural level of economic activity but a higher rate of inflation. Friedman writes, "Rising income will raise the liquidity preference . . . and the demand for loans; it may also raise prices, which will reduce the real quantity of money. These three effects will reverse the initial downward pressure on interest rates in something less than a year. Together they will tend, after . . . a year or two to return [real] interest rates to the level they would otherwise have had" (1968, 5–6).

Friedman's theory of the short-run effects of monetary policy is sometimes described as the liquidity effect theory. In recent years this theory has been the basis for a great deal of recent research, both empirical and theoretical, about the short-run effects of monetary policy.

As the discussion has indicated, Friedman's liquidity effect theory is based on the belief that in the short run the decisions of firms and households are influenced by money illusion. In this theory, an increase in production and employment occurs not because there has been a change in economic fundamentals but because a more rapid rate of monetary growth has produced a higher rate of inflation. In Friedman's words, "The monetary authority can make the market rate less than the natural rate [of interest] only by inflation" (1968, 7).

The monetary surprises/money illusion hypothesis of Friedman and Phelps seemed to reconcile classical economic principles with the existence of Phillips-type relationships (a negative relationship between inflation and the real interest rate, a positive relationship between inflation and the level of real output, and so forth) created by monetary policy. Under this hypothesis the Phillips curve continued to represent a menu of choices involving trade-offs between real and nominal variables that were available to monetary policymakers—but only in the short run.

The Accelerationist Hypothesis. In tandem with this money illusion hypothesis, monetarists held firm to the classical premise that in the long run all real economic variables such as the real interest rate or the real unemployment rate have a natural level that is determined by economic fundamentals and is completely independent of the nature of monetary policy. In their view, temporary money illusion was the only mechanism by means of which monetary policy could affect real economic activity. It followed from these premises that continuous efforts by monetary policymakers to stimulate economic activity would translate mostly into an everincreasing rate of inflation. While it might be possible for monetary policy to influence the level of interest rates (in

particular) and real economic activity (in general) in the short run, once households and firms recognized that the rate of inflation had increased, the aggregate supply would shift back and the real effects of an increased inflation rate would disappear. Further reductions in interest rates and further stimulus to economic activity could be attained only via further increases in the rate of inflation. In Friedman's words, "Let the monetary authority keep the nominal market rate for a time below the natural rate by inflation. That in turn will raise the nominal natural rate itself, once anticipations of inflation become widespread, thus requiring still more rapid inflation to hold down the market rate" (1968, 7-8). The view underlying this "accelerationist" hypothesis is that while economic decisionmakers cannot be fooled permanently by a single increase in the inflation rate, they can be fooled persistently by accelerating inflation—that is, by a price level that increases over time at an increasing rate.

The Monetarists and the Keynesians in Perspective

he monetarists' persistent attacks on the Keynesians failed to convince the Keynesians that systematic efforts to use monetary policy to affect economic activity would fail. The monetarist argument that attempts to exploit the short-run inflation-unemployment trade-off would lead to accelerating inflation convinced Keynesians that balancing the competing economic goals of keeping inflation low and keeping real economic activity brisk would be harder than they had thought. However, the argument did not convince them that this balancing act was impossible.

To reiterate, during the 1960s Keynesian theorists came to regard the Phillips curve as a menu of options between inflation and unemployment from which policy-makers could choose. They assumed that the Phillips curve was stable, which implied that monetary policy was powerful both in the short run and in the long run (that is, that money was not long-run superneutral). To Keynesians, the job of macroeconomic policymakers was to design demand-management policies that would strike the right balance between the competing problems of sustaining robust economic activity and controlling inflation.

Monetarists, on the other hand, believed the economy would be better off if the Federal Reserve supplied money according to a fixed, publicly announced formula and did not try to influence the level of real economic activity. Monetarists such as Friedman and Phelps disagreed with Keynesians regarding the effectiveness and usefulness of demand management. They viewed the "natural rate of unemployment" (the analog of Friedman's natural rate of interest: see above), together with the quantity theory of money, as solid enough arguments to assert beyond doubt the undesirability of activist mone-

tary policy and the long-run superneutrality of money. Monetarists acknowledged the possibility that monetary policy might have short-run effects on employment, interest rates, and private spending, but they believed that these effects arose exclusively from the public's misperception of the impact of changes in the price level. According to the monetarists, the only way the monetary authority could have persistent real effects was by produc-

ing an ever-accelerating rate of inflation.

The debate between the monetarists and the Keynesians sometimes took the form of disagreements about the slope of the Phillips curve. These disagreements reflected differing views about the effectiveness of monetary policy in the short run versus the long run. During the 1970s, the Keynesians attempted to capitalize on the monetarists' ten-

Friedman's liquidity effect theory is based on the belief that in the short run the decisions of firms and households are influenced by money illusion.

dency to frame the debate about monetary policy in terms of short- and long-run effects. Their strategy involved reinterpreting the Phillips curve in a way that reconciled the Keynesian and monetarist views of the timing of the inflation-unemployment relationship. This strategy forced the Keynesians to acknowledge that there were limits on the exploitability of the Philips curve.

A key element of the "compromise" offered by the Keynesians was the NAIRU, an acronym that stands for "non-accelerating inflation rate of unemployment" (see Espinosa and Russell 1997). In a diagram of the Phillips curve, the NAIRU is the unemployment rate at which the negatively sloping Phillips curve intersects Friedman's natural rate of unemployment. Monetarists believed that the existence of a natural rate implied that there was no useful trade-off between inflation and unemployment. Keynesians, however, interpreted the natural rate as a long-run constraint that policymakers have to face when trying to exploit an inflation-unemployment trade-off that remained both available and helpful in the short run. This revised Keynesian view of the trade-off was accepted by most policy-oriented economists and most economic policymakers. In the words of Tobin, the "consensus view accepted the notion of a nonaccelerating inflation rate of unemployment . . . as a policy constraint on policy" (1980, 24).

In retrospect it is clear that as much as the monetarists tended to dismiss Keynesian views, in many ways the two schools were not very far from each other—particularly in their analyses of the short-run consequences of monetary policy. These similarities become more evident when the monetarist-Keynesian debate is put in historical perspective. The years since the 1970s have witnessed the development of "neoclassical" macroeconomics—a new school of macroeconomic thought that is based on classical principles even more firmly than monetarism. One of the most influential branches of neoclassical macroeconomics is real business cycle theory. According to real business cycle pioneers such as Kydland and Prescott (1982) and Nelson and Plosser (1982), the cyclical pattern of recessions and expansions has little to do with monetary policy and can be explained almost entirely by "real shocks"—technological developments, changes in tax policy, and other unpredictable changes in economic fundamentals. Thus, the real business cycle theorists believe monetary policy has few or no effects even in the short run.

As economist Joseph Stiglitz points out, "Friedman is, in many ways, closer to the Keynesians than to the real business cycle theorists. He believes, for instance, that there are short run rigidities . . . such that any action by the monetary authority cannot immediately and costlessly be offset by changes in the price level" (1991, 48). Stated differently, the short-run predictions of the Keynesians and the monetarists differed in magnitude but not in direction. Both groups believed in a monetary policy transmission mechanism under which an increase in the money supply leads to an increase in economic activity accompanied by an increase in the general price level. The disagreement about magnitudes could, in principle, have been settled by the analysis of the empirical evidence (although in practice this was no easy task). But as long as the monetarists conceded that monetary policy had some short-run real effects it was impossible for them to make an unequivocal case against the exploitability of the Phillips curve.

In summary, the classical school saw the long-run level of economic activity as being determined independently of monetary policy. Metzler (1951) accepted much of the classical analysis but believed that there were situations in which monetary policy could have long-run real effects. The monetarists focused on money illusion as the only mechanism through which monetary policy could have real effects. In their view, economic fundamentals helped determine an individual's demand for money for transaction purposes. In the absence of surprises this money-demand relationship was fairly stable. It followed that in the long run, changes in the rate of money growth would produce proportional changes in the rate of inflation but would not affect real variables. Tobin (1965) sketched out a formal model in which changes in the rate of money growth could have long-run real effects. In his portfolio-based analysis, a permanent increase in the inflation rate led to more capital accumulation and a lower real rate of return on physical capital. Keynesians implicitly accepted the monetarist view of the role of money illusion in empowering monetary policy. They came to view Friedman's natural rates, which could be interpreted as long-run equilibrium values determined exclusively by fundamentals, as long-run constraints on policy strategies that remained effective in the short run. The short-run policy effect predictions of the Keynesians and the monetarists differed in regard to magnitude and persistence but not in regard to direction. Both schools agreed that in the short run a higher rate of money growth was associated with a higher rate of inflation, a lower real interest rate, and a spurt in economic activity. Keynesians did not themselves develop theories in which monetary policy was powerful in the short run but money was superneutral in the long run. Instead, they implicitly accepted the theoretical framework provided by their critics, the monetarists, although the two schools continued to disagree about some of the implications of this framework.

To this day, much of the economics profession continues to regard Keynesians' acceptance of the monetarists' position regarding long-run superneutrality as proof that there has been a rigorous scientific synthesis of the two theories. As discussed below, however, any synthesis of this sort is likely to be internally inconsistent.

The Neoclassical School

ealhe arguments made by Friedman and Phelps against Keynesian theory were extended by economists such as Lucas (1972) and Sargent and Wallace (1976), who became the founders of the neoclassical school.⁶ Lucas's 1972 article set the standards for neoclassical macroeconomics and, to a large extent, for all modern macroeconomics. The two pillars of his analysis were his assumption that economic decisionmakers had rational expectations and his use of a dynamic general equilibrium model. A dynamic general equilibrium model is a model that takes into account the intertemporal nature of many economic decisions and recognizes that economic variables interact with each other. Therefore, to determine the consequences of a postulated policy experiment one has to consider the relevant economic variables simultaneously and through time.

Lucas's article presented a very rigorous description of a situation in which (1) money is superneutral in the long run, and (2) the short-run real effects of monetary policy are bound to be rather limited, even in a scenario involving accelerating prices. A first step toward understanding Lucas's analysis is to recognize a key distinction between his assumptions and those of Friedman and Phelps. A simple way to describe this distinction is to say that the Friedman and Phelps analysis permitted persistent money illusion while Lucas's analysis ruled out persistent money illusion. Stated differently, the Friedman and Phelps analysis was based on the assumption that

changes in prices or wages could cause households and firms to make "bad" economic decisions—decisions they would not have made if they had used available economic information more efficiently and had displayed more flexibility in reacting to the changes. Lucas, in contrast, assumes that the public processes economic information as efficiently as possible: in particular, individuals base their current decisions on the best possible forecasts of future events. His description of this decision-making process includes specific assumptions about how people form their economic expectations.

In Lucas's model there are two types of changes in prices: temporary changes in prices in particular industries, which are caused by short-run fluctuations in the demand for the goods produced by those industries, and changes in the price level, which are caused by changes in the growth rate of the money supply. There are also two types of changes in the growth rate of the money supply: systematic, permanent changes in the average (long-run) money growth rate and unsystematic, temporary changes in the current (short-run) money growth rate. The systematic changes result from deliberate changes in policy by the central bank; they produce a permanent increase in the average rate of inflation. The unsystematic changes result from errors in the implementation of the central bank's operating procedures. They do not reflect deliberate policy decisions, and they do not affect the long-run average money growth rate or inflation rates. They do, however, produce temporary changes in the current rate of inflation.

As has been indicated, Friedman and Phelps had assumed implicitly that economic decisionmakers have access to complete economic information but fail to use it efficiently. Lucas, on the other hand, assumes explicitly that decisionmakers use any information available to them in the most efficient way but do not always have access to complete information. The particular aspect of the economy that Lucas assumes decisionmakers do not have complete information about is the relationship between changes in the relative prices of the particular goods they produce and changes in the overall price level. This information gap is important because fully informed decisionmakers will react very differently to changes in the prices of their goods that represent changes in relative prices—that is, to situations in which the prices of their goods change but the general price level remains constant, or situations in which the general price level changes but the prices of their goods change by a larger or smaller proportion—than to changes in the prices of their goods that simply follow along with changes in the overall price level. More specifically, decisionmakers have no incentive to increase their work effort and production in response to

increases in the overall price level for the same reason that one would not be any happier if a doubling of salary coincided with a doubling of the price of every good purchased. On the other hand, it makes sense for a person to increase effort and output if the relative price of the good produced has increased. Under Lucas's assumptions any such increases in effort and output will be temporary because the demand fluctuations that induce them are also temporary.

Now suppose that at a given moment in time, and in the absence of any changes in the economy's fundamen-

tals, the overall inflation rate increases because of an unsystematic increase in the money supply. As the overall inflation rate increases, prices in every sector or industry increase. However, individuals are unable to tell, immediately, whether the price increases affecting their sector are relative or absolute changes. The reason is that people are assumed to have better information about prices of the

To Keynesians, the job of macroeconomic policymakers was to design demandmanagement policies that would strike the right balance between the competing problems of sustaining robust economic activity and controlling inflation.

goods and services in their industry than about the many different prices that figure in the overall price level. This lack of complete information about the overall level of prices leads people to assume that at least part of the increase in the price of their product has been caused by an increase in its relative price. As a result, they increase their work effort and production.

The situation just described seems quite consistent with the Keynesian notion that there is a short-run trade-off between rate of inflation and the level of economic activity. But does this trade-off indicate that monetary policy is powerful, in the sense that the central bank can use it to control the level of economic activity? Is this a model of the "tightrope walk" that aggregate demand managers are often described as having to perform? If the central bank in the model can use monetary policy actions to exert continuous and repeated influence over individual decisions concerning work effort and production, then the answer to these questions may be yes.

This situation turns out not to be possible, however. Suppose that the central bank announces a permanent change in the average growth rate of the money supply. Lucas's assumption that people have rational expectations implies that they understand the nature of the relationship between money growth and inflation. As a

result, they will not increase their work effort or production in response to the resulting increase in the average rate at which prices change. Thus, permanent increases in the money growth rate have no effect on the level of output or employment, while temporary increases in the money growth rate will produce temporary increases in both output and employment.

Thus, in Lucas's model the rational expectations assumption implies that systematic changes in monetary policy should not have real effects. The rigorous nature of

To this day, much of the economics profession continues to regard Keynesians' acceptance of the monetarists' position regarding long-run superneutrality as proof that there has been a rigorous scientific synthesis of the two theories.

Lucas's analysis made his argument seem very convincing. It is important to note, however, that the argument also depends on Lucas's assumption, which is built into the structure of his model, that the effects of monetary policy on real economic activity are caused only by money illusion.

Lucas's argument can be illustrated further by returning to the context of his model and exploring its implica-

tions in a somewhat less rigorous way. Suppose that the central bank in his model attempts to exploit the apparent trade-off between inflation and output by increasing the average money growth rate without making any announcement that it has done so. It is hoping that people will make inflation-forecasting mistakes because they will not recognize that any policy change has occurred. The increase in money growth will, of course, produce a permanent increase in the average inflation rate. Initially, people will mistake this systematic, policyinduced increase in the inflation rate for an unsystematic inflation rate increase caused either by a temporary demand disturbance or by an error in the execution of the original monetary policy rule. Since they will not be sure which of these two types of unsystematic increase has occurred, their work effort and production will rise (see above). Soon, however, people will start to recognize that there is a pattern to the unusually high rates of inflation they are observing. As a result, they will begin to think it less and less likely that the next above-average increase in the inflation rate was caused by a demand disturbance, and they will begin to cut back on their above-normal production and work effort. Ultimately, they will realize that the central bank has changed policy in a way that has caused the average inflation rate to increase. At this point, the increase in the average inflation rate will no longer have any effect on work effort and production.

The scenario just described suggests that systematic changes in monetary policy may have substantial short-run effects but no long-run effects, just as the monetarists argued and just as the Keynesians ultimately conceded. Suppose, however, that the central bank tries to repeat its short-run success by changing the average inflation rate from time to time in response, say, to other changes in the state of economy. In the real world people learn from past mistakes: as a result, each time the central bank engineers another systematic change in the inflation rate people will catch on to the policy change more quickly and the effects of the change will disappear more quickly. At some point, moreover, people will figure out which events motivate the central bank to change policy; they will then be able to detect policy changes as soon as they occur. At this point the policy changes will no longer have any effects, even in the short run.

These modified rational expectations assumptions about the way people obtain and use information seem consistent with one's economic intuition about the behavior of actual households and firms. In real-life economies, most people have a very good "micro" picture of the status of their firm or industry but a fairly fuzzy "macro" picture of the state of the economy at large. However, once they start getting surprised by unexpected price changes that make their decisions work out badly they become more interested in identifying the causes of changing prices. They start to use any information available to them to try to anticipate changes in the price level and distinguish them from changes in relative prices. As a result, future price level increases have less and less surprise effect. This sort of intelligently adaptive behavior is the real-life analogue of Lucas's formal assumption that economic decisionmakers have rational expectations.

Lucas's argument, and the closely related arguments of neoclassical economists such as Sargent and Wallace (1976), left Keynesians with only two intellectually legitimate choices. First, they could have tried to capture their intuition about the effects of monetary policy in a rational expectations general equilibrium model in which money was not long-run superneutral because monetary policy derived its real effects from some source other than monetary surprises. Many economists expected this approach. Sargent, for example, writes that "in the early 70's, I thought that Modigliani, Solow, and Tobin—our heroes in those days—were missing the boat by resisting the intrusion of rational expectations into macroeconomics, instead of commandeering it. Despite the appearances of its early incarnations like Lucas's 72 JET paper, the canons of rational expectations models . . . were evidently wide enough to include Lucas's brand of monetarism or, just as readily, accommodate the completion of Tobin's criticism of monetarism by fully bringing to bear the logic of Modigliani and Miller. Modigliani, Solow, and Tobin chose not to commandeer the movement, and left it to Kareken, Wallace, Chamley, Bryant and others to draw out many of the nonmonetarist implications then waiting to be exposed." (1996, 545). In retrospect it seems clear, as this quotation indicates, that an important reason Keynesians did not pursue this strategy was because they mistakenly believed that rational expectations implied long-run superneutrality of money.

Another alternative for Keynesians might have been to concede that monetary policy was long-run superneutral but to argue that frictions of various sorts might allow monetary policy to have real effects in the short run. Taylor's work on staggered contracts (1980), his work on slow adjustment of prices (1994), and the work of Ball and Mankiw (1995) concerning "menu costs" are illustrations of this line of research. This research has faced criticisms on two fronts. First, there is little empirical evidence to support this type of nominal rigidities (see, for example, Wynne 1995). Second, there has not been a clear explanation as to why these frictions could prevent people from changing their behavior so as to offset the effects of systematic changes in monetary policy. For example, what prevents individuals from relying on mechanisms such as indexing of nominal contracts to guard against the potential negative effects of nominal rigidities?

Most Keynesians chose to ignore the neoclassical critique and the potential problems with short-term frictions. They continued to claim that monetary policy had powerful short-run effects, while accepting the monetarist critique that it was powerless in the long run. For the most part, economists outside academia—business economists and economic policymakers—have adopted this "Keynesian consensus" view. To the extent that either group of economists has attempted to justify this view, they have done so by arguing that rational expectations is a sensible assumption only in the long run. In the short run, they argued, people could and often did misread the nature and effects of changes in monetary policy.

What is wrong with the Keynesian consensus? Lucas points out that, while it may seem reasonable on its face, it suffers from serious logical problems. Com-menting on Tobin's description of the Keynesian consensus, Lucas writes, "Here we have Model A, that makes a particular prediction. We have model B, that makes a strikingly different prediction concerning the same event. The event occurs, and Model B proves more accurate. A proponent of model A concludes: 'All right, I "accept" Model B too.' Consensus economics may be a wonderful thing, but there are laws of logic which must be obeyed . . . These models gave different predictions about the same event because their underlying assumptions are mutually inconsistent. If the Friedman-Phelps assumptions are now 'accepted,' which formerly accepted Keynesian assumptions are now viewed as discarded? Tobin does not say" (1981, 560-61). Lucas goes on to spell out the monetarist (model B)-Keynesian (model A) consensus, as viewed through the Keynesian glass. He writes, "Though I refer to Tobin as 'evading' a central issue, I do not think he sees it this way at all. He writes as though he is willing to concede the 'long-run' to Phelps and Friedman [the Monetarists], claiming only the 'short-run' for Keynesians. Where is the conflict?" (561). Lucas goes on to explain that the long run consists of a sequence of short runs. If a policymaker conducts short-run policy by choosing an annual money growth rate based on model

A (the Keynesian model) every year, then he or she has implicitly used model A to pick the average rate of money growth for the long run. It is logically inconsistent to pretend that the long-run average money growth rate using model B (the monetarist model) can be a guide. Suppose, for example, that the central bank decides that in the long run the optimal growth rate of the money supply is 5 percent per

Lucas's 1972 article set the standards for neoclassical macroeconomics and, to a large extent, for all modern macroeconomics.

year. However, it decides on the basis of short-run considerations that it would be a good idea to increase the money growth rate to 6 percent for the coming year. The same thing happens again in the following year, and in the year after, and so on. The end result is a departure from the optimal long-run money growth rate that may have adverse consequences for the economy. Thus, Lucas observes that "if we concede that Model A gives us an inaccurate view of the 'long-run,' then we have conceded that it leads us to bad short-run situations as well" (560–61).

Monetary Policy after Lucas. Starting in the late 1960s, Keynesian economic theory was the victim of a succession of setbacks, including the monetarist critique of Friedman and Phelps, the combination of high inflation and high unemployment that the United States experienced during the 1970s, and the neoclassical critique of Lucas (1972) and Sargent and Wallace (1976). As Keynesian theory lost ground in the academic community, so did belief in the power of monetary policy. In fact, much of the early work by neoclassical economists followed Lucas (1972) by constructing models that made debating points against the Keynesians by demonstrating that systematic changes in monetary policy would have no real effects, even in the short run. Unsystematic policy actions might have a short-lived influence on the level of economic activity, but any attempt to use systematic changes in policy to exploit this influence would be frustrated by changes in the expectations of the public.

The fact that the model described by Lucas (1972) became the "industry standard" in neoclassical theory has encouraged other neoclassical economists to focus on models that display long-run superneutrality of money. In recent years, the most popular vehicle for research on monetary policy by neoclassical economists has been the real business cycle model. In this model money is long-run superneutral, but temporary changes in monetary policy can generate small "liquidity effects" of the sort described in Friedman (1968). (See, for example, Lucas 1990; Christiano and Eichenbaum 1991, 1992; Fuerst 1992; Dow 1995).

Lucas's article presented a very rigorous description of a situation in which (1) money is superneutral in the long run, and (2) the short-run real effects of monetary policy are bound to be rather limited, even in a scenario involving accelerating prices.

During the mid-1970s economist Harry Johnson, reviewing what he labeled the Keynesian revolution and the monetarist counterrevolution, commented that "the monetarist counterrevolution has served a useful purpose, in challenging and disposing of a great deal of the intellectual nonsense that accumulates after a successful ideological revolution . . . If we are lucky, we shall

be forced as a result of the counterrevolution to be both more conscious of monetary influences on the economy and more careful in our assessment of their importance" (1975, 106).

In fact, the monetarist counterrevolution had mixed effects on economists' views concerning the importance of monetary policy. On the one hand, monetarist arguments convinced many Keynesians that monetary policy had many of the same powers that they had attributed to fiscal policy. On the other hand the monetarists, as has been pointed out, completely dismissed the possibility that monetary policy might have long-run real effects. To the extent that Keynesians conceded this point they were also conceding that the importance of monetary policy was quite limited.

As shown above, the period of the monetarist counterrevolution was also a period when a few economists began to try to identify explicit mechanisms that would allow monetary policy to have long-run real effects. Metzler (1951) and Tobin (1965) developed theories that allowed the Keynesian, conventional wisdom to be extended to the long run. These theories allowed permanent increases in the money supply growth and inflation rates to be causally associated with permanently lower real interest rates and permanently higher levels of output.

Lucas's (1972) work suggested that macroeconomic theories of all sorts were in need of reevaluation. The theories of Metzler, Tobin, and the ones derived from Phillips's analysis were no exception. Lucas's interpretation of the Phillips curve analysis has been described above. The next section reviews subsequent research that tries to reformulate Metzler's and Tobin's theories using the neoclassical methodology Lucas introduced. This research indicates that departures from long-run superneutrality are possible because monetary policy does not necessarily derive all (or any) of its power from money illusion. Instead, changes in monetary policy may have lasting effects because it affects the supply of or the demand for credit.

Some Neoclassical Models That Deliver Long-Run Nonsuperneutralities

his section looks at the three challenges facing economists who want to develop neoclassical models that deliver results similar to those of Metzler, Phillips, and Tobin. The first challenge is simply to construct a plausible neoclassical model in which money is not long-run superneutral. The second challenge is to identify a mechanism under which the departures from superneutrality work in the "right direction," that is, a mechanism that allows increases in the money supply growth rate to be causally associated with lower real interest rates and higher levels of output. The third challenge is to find a mechanism that has some hope of generating departures from superneutrality that are large enough to have practical importance.

The Tobin Effect. An answer to the first challenge is to rely on the credit market as the ultimate source of the real effects of monetary policy. In this respect one could follow Tobin (1965). Tobin's analysis is based on the idea that the increase in the inflation rate that is induced by an increase in the money supply growth rate increases the supply of credit at any real interest rate. It does so because when the inflation rate rises money becomes a relatively unattractive asset, and the public wishes to cut back on its money balances and increase its holdings of bank accounts, bonds, stock, and other financial assets. Thus, there is a decrease in the demand for money and a matching increase in the supply of credit. The Tobin effect mechanism allows a permanent easing of monetary policy (a higher money growth rate) to lead to a higher inflation rate, a lower real interest rate (due to the increased supply of credit), and a higher level of output (due mostly to an increase in the capital stock).

Many economists believe that financial intermediation is one of the most important channels through which changes in monetary policy affect the economy (see for instance Bernanke and Gertler 1995). The central bank may be able to affect the composition of financial intermediaries' portfolios without relying on monetary sur-

prises. Thus, permanent changes in monetary policy may affect financial intermediaries in a fundamental way and may have long-run real effects. It follows that a natural environment in which to analyze the Tobin effect would be one in which financial intermediaries were explicitly developed.

The starting point for assessing this possibility should be a realistic model of financial intermediation. Bencivenga and Smith (1991) were among the first economists to include financial intermediaries in a dynamic general equilibrium macroeconomic model. The Bencivenga-Smith intermediaries are similar to actual intermediaries in accepting deposits from, and lending to, a large number of individuals. They are also similar to actual intermediaries in making loans that are less liquid than the deposits they accept. As a result, they are forced to hold a liquid asset (money) on reserve to cover sudden deposit withdrawals.

In the Bencivenga-Smith model individuals could, in principle, manage their own asset portfolios (as in Tobin 1965). However, the financial intermediaries have an actuarial advantage over individuals in structuring a portfolio. Consequently, under most circumstances people will prefer to delegate this activity to financial intermediaries. Although Bencivenga and Smith's work contains the elements needed to pursue an analysis of the long-run effects of permanent changes in monetary policy, their analysis concentrates on the long-run implications of financial intermediaries for an economy's long-run performance. Based on the Bencivenga and Smith model, Espinosa and Yip (1998) study the growthinflation implications of alternative fiscal and monetary policies in the presence of financial intermediaries. Espinosa and Yip can, thus, draw some qualitative lessons on the Tobin effect in a dynamic general equilibrium model that explicitly models financial intermediaries. Before listing their findings, it is useful to briefly review some recent empirical results on the relationship between inflation and growth.

Inflation and Growth. In part because money has been assumed to be long-run superneutral, there has not been much research on the long-run relationship between inflation and growth. Recently, however, interest in theoretical and empirical analysis of this relationship has revived. The empirical findings are not always in agreement. DeGregorio (1992) and Barro (1995) uncover a significant negative correlation between inflation and economic growth. On the other hand, Bullard and Keating (1995) and Bruno and Easterly (1998) do not find strong support for such an inverse relationship. Bullard and Keating find that the direction of the growthinflation relationship depends crucially on the initial level of the inflation rate. In countries in which the rate of inflation starts out relatively low, a permanent increase in the inflation rate actually increases the longrun level of economic activity. Only for countries with relatively high initial inflation rates do Bullard and Keating find that permanent increases in the rate of inflation negatively affect long-run growth. These findings are partly confirmed by Bruno and Easterly, who are able to find an inverse relationship between inflation and growth only when the rate of inflation exceeds some critical value.

Clearly, these empirical studies do not settle whether monetary policy can have real effects that do not spring from monetary surprises and whether these effects are likely to be of the type described by Tobin, with higher inflation being associated with higher rates of growth.

Espinosa and Yip (1998) address these questions in a model based on the model developed by Bencivenga and Smith (1991). Their analysis emphasizes the point (to be made very explicitly below) that fiscal and monetary policy are inevitably linked by the government budget constraint. In their model, monetary policy consists of changes in the growth rate of the money supply that are necessitated by

Keynesians mistakenly believed that rational expectations implied long-run superneutrality of money.

changes in fiscal policy—specifically, by changes in the government budget deficit as a fraction of GDP.

Espinosa and Yip show that their model can produce the positive long-run relationship between inflation and growth that was predicted by Tobin. However, it is also possible for the model to produce situations in which lower rates of inflation result in higher rates of growth. The direction of the inflation-growth relationship depends on, among other things, the initial inflation rate, the degree of risk aversion of the average depositor, and the size of the government budget deficit. Thus, the Espinosa-Yip analysis provides a theoretical framework that helps reconcile the conflicting empirical findings about the direction of the long-run relationship between inflation and growth that were described in the preceding subsection.

Fiscal Policy and Open Market Operations. The Tobin effect has a potential drawback as a theory of the real effects of monetary policy (see, for example, Danthine, Donaldson, and Smith 1987). The shift in the credit supply curve produced by an increase in the inflation rate is essentially equal to the reduction in money demand that the increased inflation induces. Money demand is quite small (a small fraction of total output, or total assets, and so forth) and statistical evidence (for example, Hoffman and Raasche 1991) suggests that it is not very sensitive to changes in the inflation rate. As a

result, the Tobin effect of moderate changes in the inflation rate on real interest rates and output is likely to be small.⁷

An alternative mechanism for linking monetary policy and the supply of credit has been developed by Sargent and Wallace (1981). This mechanism is based on the fact that changes in monetary policy affect the government's stream of revenues and thus necessitate changes in fiscal policy. To gain a better understanding of this mechanism, it is useful to review some basic elements of a government's budget constraint.

An important premise of the research described in this section is that both fiscal and monetary policy actions are constrained by the government's need to finance its expenditures. Consequently, these two types of government policy cannot be devised or executed independently from each other. The government of a country must decide on the level of government spending to finance domestically, how much of its domestic financing will rely on current taxes, and how much will take the form of newly issued debt. Stated differently, it is the government budget deficit that determines the need for new issues of government debt. Since government borrowing competes with private borrowing in the credit market, the amount of government borrowing is likely to influence the level of real interest rates.

Government policy concerning taxes, debt, and deficits is usually described as fiscal policy. The analysis just presented suggests that fiscal policy may affect real interest rates. However, monetary policy has a fiscal policy aspect to it because it may play a role in determining the size of the government budget deficit. To the extent that monetary policy has this effect, this analysis suggests that it will also have an impact on real interest rates. Thus, monetary policy may influence the real economy in ways that do not involve inflation surprises. If this influence can persist in the long run then money may not be long-run superneutral.

In practice, monetary policy is carried out via open market operations. Open market operations produce changes in the composition of the government's portfolio of liabilities—debt (bonds and bills) versus money.8 Given the amount of government bonds currently outstanding, the government must decide what fraction of these bonds (if any) it will "monetize" by purchasing them with newly created currency. This decision, which determines the composition of the government's liability portfolio, also determines the amount of outstanding government debt in the credit markets and consequently has an impact on the market real rate of interest. More specifically, changes in the growth rate of the money supply affect the volume of government revenue from currency seigniorage (the "inflation tax"). Sargent and Wallace (1981) assume that the government's primary (net of interest) budget deficit is fixed by the tax and

spending decisions of Congress and is not affected by changes in monetary policy. Consequently, the only way the government can offset the changes in its revenues that are caused by changes in monetary policy is to change the size of the national debt. Thus, this mechanism can be thought of as the neoclassical successor of Metzler (1951) (because of Metzler's emphasis on open market operations as the mechanism through which non-long-run superneutrality results could be attained).

The size of the national debt has substantial effects on the state of the government budget. On the one hand, the government has to pay interest on the debt. On the other hand, as the economy grows the government can allow the national debt to grow at the same rate without increasing the size of the debt relative to the economy. The relationship between these two factors determines whether debt service is a financial burden for the government or whether the existence of the national debt actually increases the amount of government revenue.

To see why the latter situation is possible, suppose the government borrows just enough each year to keep the debt-GDP ratio constant. If the economy is growing, it will increase its borrowing each year by an amount that causes the real national debt to grow at the same rate as real GDP. Although the government will have to use some of the proceeds of this new borrowing to pay the interest on the current debt, if the real (also inflationadjusted) interest rate on the debt is lower than the real GDP growth rate then the government will have funds left over to use for other purposes. In this case, the national debt actually provides the government with revenue on net. This source of revenue is sometimes referred to as bond seigniorage. 10 The difference between the real growth rate and the real interest rate is the net real amount that each real dollar of debt contributes to the government budget each year.

If the real interest rate on the government debt is higher than the output growth rate then the government's new borrowing will not be enough to cover the interest on the existing debt. As a result, some of this interest will have to be covered by funds from other sources. In this case the national debt is a financial burden for the government. (One can think of this as a case in which bond seigniorage revenue is negative.) The difference between the real interest rate and the real growth rate is the net real amount that each real dollar of debt costs the government each year.¹¹

Once it is known whether the national debt is a source or a use of government funds one is in a position to determine how a change in the size of the national debt will affect the government's budget position. Other things being equal, an increase in government borrowing that increases the size of the national debt represents an increase in the quantity of credit demanded at each real rate of interest and will consequently produce an

increase in the real interest rate. If the real interest rate is relatively high, so that the debt is a burden on the government budget, then the combination of a larger debt and a larger unit cost of financing the debt means that the debt will definitely become costlier to the government. Conversely, a smaller debt that will result in a lower real interest rate will reduce the government's costs. As a result, when the government cuts the money supply growth and inflation rates and loses money from currency seigniorage, it must compensate by cutting back on its borrowing and driving the real interest rate down. As a result, tighter monetary policy produces lower real interest rates and a higher level of output.

Suppose, on the other hand, that the real interest rate is relatively low, so that the national debt is a source of revenue for the government. In this case, a given change in the size of the debt (say, an increase) can either increase or decrease government revenue from bond seigniorage. An increase in the size of the debt tends to cause bond seigniorage revenue to increase: this is the "tax base effect" of the increase. On the other hand, an increase in the government debt drives the real interest rate closer to the output growth rate and reduces the real seigniorage revenue produced by each real dollar of debt. This is the "tax rate effect" of the increase. If the tax base effect is stronger than the tax rate effect then an increase in the size of the debt will increase the government's bond seigniorage revenue; otherwise, the amount of revenue will fall.

The tax rate effect tends to be largest when the government debt is large, because in this case any change in the real interest rate affects the revenue produced by a large volume of debt. Conversely, the tax base effect tends to be largest when the real interest rate is low because each dollar of debt generates a lot of revenue. A low real interest rate tends to be associated with a small volume of government debt, since when the real interest rate is low private credit demand is high and private debt crowds out government debt. Conversely, a high real interest rate tends to be associated with a large government debt. As a result, when the real interest rate is relatively low—well below the output growth rate—an increase in the size of the national debt tends to increase bond seigniorage revenue while when the real interest rate is higher an increase in the size of the debt tends to decrease the amount of revenue.

One can now put all the pieces of this story together to determine the possibilities for the long-run real effects of monetary policy. If the real interest rate is higher than the output growth rate, or lower than the output growth rate but not too much lower, then an increase in the size of the national debt decreases government bond seigniorage revenue and vice-versa. Thus, a decrease in the money growth and inflation rates that reduces government revenue from currency seigniorage will force the government to reduce

the size of its debt and will drive the real interest rate down. This is the scenario described by Wallace (1984); it has the implication that monetary tightening will increase the level of real GDP in the long run. On the other hand, if the real interest rate is substantially below the output growth rate then a decrease in the money growth and inflation rates will allow the government to increase

The proposition that monetary policy does not have long-run real effects is far from unequivocally established.

This is the scenario described by Espinosa and Russell (1998a, b). It is similar to the Tobin effect in having the Keynesian, or conventional, implication that a monetary tightening will reduce the level of real output.

Historically, the average real interest rate on U.S. government debt has been well below the average U.S. output growth rate. This situation makes Espinosa and Russell's Keynesian scenario seem plausible empirically. An additional reason why the scenario is appealing is that it weakens the link between the size of the money supply and the size of the shift in the credit supply curve that is induced by a change in monetary policy—the link that keeps the Tobin effect small. Although the fact that the money supply is small relative to GDP means that a change in the inflation rate will have a relatively small impact on government revenue (from currency seigniorage), if it takes a relatively large increase in the real interest rate to produce a substantial decrease in government revenue from bond

^{7.} Of course, if fiscal and monetary policy interactions led not only to long-run output level changes but to output growth changes, the Tobin effect could be of more significance.

^{8.} In principle, of course, there exists the possibility that such a swap of liabilities results in no effects either real or nominal, either in the short or long term (a case made by Wallace 1984 and Sargent and Smith 1987 but not reviewed here), but under most circumstances it will.

^{9.} The debt-GDP ratio cannot continue to grow forever. Otherwise, at some point the debt would get so large relative to households' income that it would be impossible for them to save enough to hold it.

^{10.} This term seems to have been first used by Miller and Sargent (1984).

^{11.} Thus, if the real interest rate is 2 percent higher than the real growth rate then each dollar of debt costs the government two cents each year, adjusted for inflation.

seigniorage then the resulting change in the real interest rate and the level of output could still be large.

Why might it take a large change in the real interest rate to produce a substantial change in the revenue from bond seigniorage? When the real interest rate is low, the tax rate effect and the tax base effect tend to work against each other. As a result, the net change in the amount of rev-

The possibility that monetary policy has substantial long-run real effects deserves more attention from economists and policymakers. enue that is produced by a change in the real interest rate can be quite small. In fact, there is always a range of real interest rates over which the two effects offset each other almost perfectly. Over this range, the ratio of the change in the real interest rate to the change in the amount of revenue it produces will be extremely large.

The bottom line here is that, at least in

principle, the Espinosa-Russell variant of the Sargent-Wallace "unpleasant arithmetic" can give us just what is needed: a theory that explains how a moderate but permanent increase in the money supply growth and inflation rates might result in a fairly large decrease in the real interest rate and a fairly large increase in the level of output.

Before concluding this section it is important to emphasize that the research just reviewed composes a relatively small part of the growing academic literature that studies the long-term effects of monetary policy in neoclassical models. Related work in this area includes Haslag (1998), Bhattacharya and others (1997), Schreft and Smith (1997), and Bullard and Russell (1998a, b). One implication of this line of research is that monetary economists may have spent too much time trying to forge direct links between changes in monetary policy and changes in the unemployment rate and the output growth rate. Instead, they perhaps should be devoting more effort to understanding the relationship between monetary policy and the economic fundamentals that drive saving and production decisions and also to exploring the relationship between monetary policy variables and "real" macroeconomic variables such as the government deficit, real interest rates, reserve requirements, and other variables that link the money market to the credit market.

Conclusion

his article has reviewed the history of the view that monetary policy has real effects in the short run but no such effects in the long run (so that money is long-run superneutral). This view grew out of a debate

between the adherents of two influential schools of macroeconomic thought, the monetarists and the Keynesians. The final result of this conflict was a unilateral, Keynesian-produced synthesis that developed during the 1970s. Under this synthesis the Keynesians accepted the monetarists' view that money was superneutral in the long run but continued to disagree with them about the magnitude and desirability of the short-run effects of monetary policy on real interest rates, real GDP, unemployment, and other real variables.

The article has argued that the beliefs that monetary policy is powerful in the short run and that money is superneutral in the long run may not be mutually consistent. The basic problem with most theories that reconcile these beliefs is that they rely directly or indirectly on the assumption that economic decisionmakers are victims of money illusion. If money illusion is the reason monetary policy has real effects, however, then its short-run real effects will be small and policymakers will not be able to exploit them systematically to achieve their goals. This point has been demonstrated in seminal work by Lucas (1972).

In the years since the 1970s, academic macroeconomics has slowly but surely embraced the neoclassical methodology pioneered by Lucas, which employs dynamic general equilibrium models and assumes that decision-makers have rational expectations. The results of Lucas's work and that of a number of other neoclassical economists has served to further strengthen the monetarist position concerning long-run superneutrality of money.

In recent years, empirical studies of the impact of monetary policy have concentrated on identifying its short-run effects. This focus has been motivated, at least in part, by the conviction that money is long-run superneutral. Many researchers seem to believe that there is overwhelming empirical evidence in favor of long-run superneutrality. In reality, however, the proposition that monetary policy does not have long-run real effects is far from unequivocally established: indeed, an exploration of the empirical literature on long-run superneutrality could easily be the subject of a separate article. For the purposes of this article, it may suffice to cite a remark by Robert King and Mark Watson, two prominent macroeconomists whose empirical research has produced evidence both for and against long-run superneutrality. King and Watson (1992) report that for the United States during the postwar period the data do not appear to be consistent with the hypothesis that, over the long run, money is superneutral or that nominal interest rates move one-for-one with inflation.

The fact that the empirical evidence on the long-run superneutrality of monetary policy is not as overwhelming as some analysts believe suggests that there may be a need to look at theories that explore potential sources of long-run real effects for monetary policy. As this article

has explained, Lucas's path-breaking work was an attempt to conduct a rigorous analysis of the logical consequences of the monetarist assumption that the real effects of monetary policy result from monetary surprises—a fact that has led Tobin, a leading Keynesian, to refer to Lucas's model as the "Monetarist Mark II" model. Lucas did not attempt to argue that every reasonable combination of assumptions would produce superneutrality, and it is consequently a mistake—albeit a very common mistake, even in the academic community—to equate neoclassical economics with the proposition that money is superneutral.

To repeat, Lucas's renowned 1972 paper employed innovative methodology to explore the implications of a very particular set of assumptions. The methodology is logically separate from the assumptions and can be used to analyze the consequences of very different assumptions. In fact, it is possible that monetary policy influences real economic activity for reasons completely different from the ones Lucas identified. In a recent interview in

New Yorker magazine, Lucas acknowledges that the real effects of monetary policy may not result from unexpected policy changes. He comments, "Monetary shocks just aren't that important. . . . There's no question, that's a retreat in my views" (Cassidy 1996, 55).

Abandoning the assumption that policy surprises are the main reason monetary policy can have real effects leaves two options. One is to accept the view of the real business cycle theorists that Federal Reserve policy actions are essentially irrelevant. A second option is to attempt to identify alternative channels through which the monetary authority could affect real economic activity. This article has reviewed a small part of the recent academic literature that explores the second option. The results reported in this literature indicate that monetary policy may be a great deal more powerful than most academic economists believe. They also suggest that the possibility that monetary policy has substantial long-run real effects deserves more attention from economists and policymakers.

REFERENCES

Ball, Laurence, and N. Gregory Mankiw. 1995. "Relative-Price Changes as Aggregate Supply Shocks." *Quarterly Journal of Economics* (February): 161–93.

BARRO, ROBERT J. 1995. "Inflation and Economic Growth." Bank of England *Quarterly Bulletin* (May): 166–76.

Bencivenga, Valerie R., and Bruce D. Smith. 1991. "Financial Intermediation and Endogenous Growth." *Review of Economic Studies* 58:195–209.

Bernanke, Ben S., and Mark Gertler. 1995. "Inside the Black Box: The Credit Channel of Monetary Policy Transmission." NBER Working Paper No. 5146.

Bhattacharya, Joydeep, Mark G. Guzman, Elisabeth Huybens, and Bruce D. Smith. 1997. "Monetary, Fiscal, and Reserve Requirement Policy in a Simple Monetary Growth Model." *International Economic Review* 38:321–50.

Bruno, Michael, and William Easterly. 1998. "Inflation Crises and Long-Run Growth." *Journal of Monetary Economics*, forthcoming.

BULLARD, JAMES, AND JOHN KEATING. 1995. "The Long-Run Relationship between Inflation and Output in Postwar Economies." *Journal of Monetary Economics* 36:477–96.

BULLARD, JAMES, AND STEVEN RUSSELL. 1998a. "How Costly Is Sustained Low Inflation for the U.S. Economy?" Federal Reserve Bank of St. Louis and IUPUI Working Paper.

——... 1998b. "An Empirically Plausible Model of Low Real Interest Rates and Unbacked Government Debt." *Journal of Monetary Economics*, forthcoming.

Cassidy, John. 1996. "The Decline of Economics." *The New Yorker*. December 2, 50–60.

Christiano, Lawrence, and Martin S. Eichenbaum. 1991. "Liquidity Effects, Monetary Policy, and the Business Cycle." Federal Reserve Bank of Minneapolis and Northwestern University Working Paper.

Danthine, Jean-Pierre, John B. Donaldson, and Lance Smith. 1987. "On the Superneutrality of Money in a Stochastic Dynamic Macroeconomic Model." *Journal of Monetary Economics* 20:475–99.

DEGREGORIO, JOSE. 1992. "The Effects of Inflation on Economic Growth." European Economic Review 36:417–24.

Dow, James P., Jr. 1995. "The Demand and Liquidity Effects of Monetary Shocks." *Journal of Monetary Economics* 36 (December): 91–115.

ESPINOSA-VEGA, MARCO, AND STEVEN RUSSELL. 1997. "History and Theory of the NAIRU: A Critical Review." Federal Reserve Bank of Atlanta *Economic Review* 82 (Second Quarter): 4–25.

———. 1998a. "Can Higher Inflation Reduce Real Interest Rates in the Long Run?" *Canadian Journal of Economics* 31:92–103.

———. 1998b. "The Long-Run Real Effects of Monetary Policy: Keynesian Predictions from a Neoclassical Model." Federal Reserve Bank of Atlanta Working Paper 98-6, April.

ESPINOSA-VEGA, MARCO, AND CHONG YIP. 1998. "Fiscal and Monetary Interactions in an Endogenous Growth Model with Financial Intermediaries." *International Economic Review*, forthcoming.

FISHER, IRVING. 1926. The Purchasing Power of Money: Its Determination and Relation to Credit Interest and Crises. New York: Macmillan Company.

FRIEDMAN, MILTON. 1956. "The Quantity Theory of Money—A Restatement." In *Studies in the Quantity Theory of Money*, edited by Milton Friedman. Chicago: University of Chicago

FUERST, TIMOTHY J. 1992. "Liquidity, Loanable Funds, and Real Activity." *Journal of Monetary Economics* 29:3–24.

GREENSPAN, ALAN. 1997. "Monetary Policy Testimony and Report to the Congress." Humphrey-Hawkins Testimony. Available on-line at http://www.bog.frb.fed.us/boarddocs/HH/9707Test.htm [August 12, 1998].

Haslag, Joseph. 1998. "Monetary Policy, Banking and Growth." *Economic Inquiry* 36:489–500.

HOFFMAN, DENNIS L., AND ROBERT H. RAASCHE. 1991. "Long-Run Income and Interest Elasticities of Money Demand in the United States." *Review of Economics and Statistics* 73:665–74.

JOHNSON, HARRY G. 1975. On Economics and Society. Chicago: University of Chicago Press.

KEYNES, JOHN M. 1964. *The General Theory of Employment, Interest, and Money.* San Diego: Harcourt Brace and Company. Original edition. 1953.

King, Robert, and Mark W. Watson. 1992. "Testing Long Run Neutrality." National Bureau of Economic Research Working Paper No. 4156.

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

LUCAS, ROBERT E. 1972. "Expectations and the Neutrality of Money." *Journal of Economic Theory* 4:103–24.

METZLER, LLOYD A. 1951. "Wealth, Saving, and the Rate of Interest." *Journal of Political Economy* 59 (April): 93–116.

MEYER, LAURENCE H. 1997. "Monetary Policy Objectives and Strategy." *Business Economics* 32 (January): 17–20.

MILLER, PRESTON J., AND THOMAS J. SARGENT. 1984. "A Reply to Darby." Federal Reserve Bank of Minneapolis *Quarterly Review* (Fall): 21–26.

Nelson, Charles R., and Charles I. Plosser. 1982. "Trends and Random Walks in Macroeconomic Time Series." *Journal of Monetary Economics* 10:139–62.

PHELPS, EDMUND. 1967. "Phillips Curves, Expectations of Inflation, and Optimal Unemployment over Time." *Economica* 34 (August): 254–81.

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

Sargent, Thomas J. 1996. "Expectations and the Nonneutrality of Lucas." *Journal of Monetary Economics* 37:535–48.

SARGENT, THOMAS J., AND BRUCE SMITH. 1987. "Irrelevance of Open Market Operations in Some Economies with Government Currency Being Dominated in Rate of Return." *American Economic Review* 77 (March): 78–92.

Sargent, Thomas J., and Neil Wallace. 1976. "Rational Expectations and the Theory of Economic Policy." *Journal of Monetary Economics* 2:169–83.

——. 1981. "Some Unpleasant Monetarist Arithmetic." Federal Reserve Bank of Minneapolis *Quarterly Review* (Fall): 1–17.

Schreft, Stacey L., and Bruce D. Smith. 1997. "Money, Banking, and Capital Formation." *Journal of Economic Theory* 73:157–82.

STIGLITZ, JOSEPH E. 1991. "Alternative Approaches to Macroeconomics: Methodological Issues and the New Keynesian Economics." NBER Working Paper No. 3580.

TAYLOR, JOHN B. 1980. "Aggregate Dynamics and Staggered Contracts." *Journal of Political Economy* 88, no.1:1–23.

——. 1994. "The Inflation/Output Variability Trade-off Revisited." In *Goals, Guidelines, and Constraints Facing Monetary Policymakers.* Conference Series No. 38. Boston: Federal Reserve Bank of Boston.

TOBIN, JAMES. 1965. "Money and Economic Growth." *Econometrica* 33:671–84.

Wallace, Neil. 1984. "Some of the Choices for Monetary Policy." Federal Reserve Bank of Minneapolis *Quarterly Review* (Winter): 15–24.

Wynne, Mark A. 1995. "Sticky Prices: What Is the Evidence?" Federal Reserve Bank of Dallas *Economic Review* (First Quarter): 1–12.