

# THE QUANTITY THEORY OF MONEY: ITS HISTORICAL EVOLUTION AND ROLE IN POLICY DEBATES

One of the oldest surviving economic doctrines is the quantity theory of money, which in its simplest and crudest form states that changes in the general level of commodity prices are determined primarily by changes in the quantity of money in circulation. This theory dates back at least to the mid-16th century when the French social philosopher Jean Bodin first attributed the price inflation then raging in Western Europe to the abundance of monetary metals imported from the mines of the Spanish colonies in South America. After undergoing considerable refinement, elaboration, amendment, and extension in the late 17th and 18th centuries by John Locke, Richard Cantillon, and David Hume, the quantity theory was integrated into the mainstream of orthodox monetary tradition. Forming the central core of 19th century classical monetary analysis, the quantity theory provided both the dominant conceptual framework for the interpretation of financial events in that century and the chief intellectual foundations of orthodox policy prescriptions designed to preserve the gold standard. Today the quantity theory survives and flourishes in the doctrines of the so-called *monetarist* school emanating from such institutions as the University of Chicago and the Federal Reserve Bank of St. Louis. Spearheaded by economists Milton Friedman, Karl Brunner, Allen Meltzer, Philip Cagan, and others, contemporary monetarists continue to expound quantity theory propositions similar to those enunciated by their classical predecessors.

The quantity theory has not gone unchallenged, however. As controversial as it is ancient, the quantity theory has probably stimulated more debate than any other single topic in the field of monetary theory. Some of the leading monetary controversies of the past two centuries, including the Bullionist and Currency School-Banking School debates of the 1800's, and the controversy between Keynes and the neo-classical economists in the 1930's, have revolved around issues relating to the quantity theory. Moreover, the debate shows little sign of subsiding. Many of the same quantity theory-related issues appearing in the earlier debates figure prominently in the current controversy between the monetarist and post-Keynesian schools of thought.

The purpose of this article is to introduce ECONOMIC REVIEW readers to this much-debated theory

and to the major monetary controversies surrounding it. Accordingly, the article (1) examines the content and implications of the key propositions of the theory, (2) traces the evolution of these propositions from their 17th and 18th century origins to their present embodiment in monetarist doctrine, (3) sketches the role played by the quantity theory in the Bullionist, Currency-Banking School, and Keynesian policy debates, and (4) outlines the major criticisms leveled against the theory during the past two centuries.

## WHAT IS THE QUANTITY THEORY?

Essentially the quantity theory is a hypothesis about the main cause of changes in the value or purchasing power of money. According to the theory, changes in the value of money are determined chiefly by changes in the quantity in circulation. When money becomes abundant, its value or purchasing power falls, and consequently the average of commodity prices rises. Conversely, if money becomes scarce, its purchasing power rises and general prices fall. In short, the quantity theory states that the stock of money ( $M$ ) is the main determinant of the price level ( $P$ ).

This brief statement of the theory, however, does not do it full justice. More than just the bare conclusion that money governs prices, the theory also consists of a set of interrelated propositions or postulates that support that conclusion. The most important of these propositions refer to (1) the proportionality of  $M$  and  $P$ , (2) the active or causal role of  $M$  in the monetary transmission mechanism, (3) the neutrality of money, (4) the monetary theory of the price level, and (5) the exogeneity of the nominal stock of money.

**The Proportionality Postulate** The first proposition states that  $P$  will vary in exact proportion to changes in the quantity of  $M$ , i.e., a given percentage change in the stock of money will result in an identical percentage change in commodity prices. Associated with the strict classical version of the quantity theory, this proposition follows from the assumption that people want to hold for transactions purposes a constant quantity of real (price-deflated) cash balances,  $M/P$ , at the economy's full-capacity level of real output. Because these cashholders look to the

purchasing power rather than to the mere money value of their cash balances, the price level must vary in direct proportion to the nominal money supply to maintain real balances intact.

It should be noted that the proportionality postulate implies that the demand for real cash balances and its counterpart, the circulation velocity of money, are completely stable. For if velocity or the demand for money were unstable, i.e., subject to erratic and unpredictable shifts, one could not safely predict that  $M$  and  $P$  would change in the same ratio. Suppose, for example, that a doubling of the nominal money supply,  $M$ , is accompanied by (1) a halving or (2) a doubling of the desire for real balances. The price level would have to quadruple in the first case and remain unchanged in the second if real balances,  $M/P$ , were to attain their desired levels. In fact, any arbitrary shift in the amount of real balances demanded would necessitate a percentage change in  $P$  different from that of  $M$ . Only if the demand for real balances remains unchanged will the proportionality relationship hold. It follows, therefore, that the strict version of the quantity theory must assume complete stability of the demand for money if it is to predict that money and prices will show equiproportionate variations.

**Causal Role of Money** A second key proposition of the quantity theory states that the direction of causation or channel of influence runs from  $M$  to  $P$ , i.e., monetary changes precede and cause price level changes. In this cause-and-effect relationship, money is seen as the active variable and the price level as the passive or dependent variable.

One important implication stemming from the interpretation of money as the active variable is that the proportionality between money and prices refers to an equilibrium condition established via a dynamic adjustment process, and not to an identity that holds at all points in time. The lead-lag, cause-effect relationship between money and prices implies that a change in  $M$  initially creates a disequilibrium between  $M$  and  $P$ . This disequilibrium then invokes forces that cause  $P$  to change. Prices continue to change until proportionality is restored and the disequilibrium is eliminated.

For such an adjustment process to occur, however, there must be some mechanism, channel, or linkage through which monetary impulses are transmitted to the price level. Traditionally, two main transmission mechanisms have been identified, namely, the (1) *direct expenditure* and (2) *indirect interest rate* mechanisms. The *direct* mechanism refers to the process by which the impact of a monetary change is channeled to the price level via a prior effect on the

demand for goods. The key link in this process is the relationship between the rate of spending on the one hand and the discrepancy between actual and desired real balances on the other. Variations in the rate of spending are seen as the means by which actual real cash balances are adjusted to the level that people desire to hold. Thus, for example, starting from a position of monetary equilibrium, an increase in the money supply initially will raise real cash balances above the pre-existing desired level. Cashholders will be left with more money than they want to hold, thereby prompting them to get rid of the excess via spending for goods. Given that the economy is operating at full capacity, however, the increased spending will exert upward pressure on prices. Spending, prices, and nominal income will continue to rise until cashholders are just satisfied to hold the nominal money in existence. Equilibrium is restored when  $P$  has risen sufficiently to bring real cash balances back to the desired level. In brief, the direct mechanism relies on the disequilibrium between actual and desired real balances to induce the spending that ultimately causes prices to change in proportion to the monetary injection. The sequence runs directly from money to spending to prices.

By contrast, the *indirect* mechanism refers to the process by which a monetary change influences spending and prices indirectly via its prior effect on the interest rate. In this process, a monetary injection first causes the rate of interest to fall, thereby stimulating business investment spending and thus exerting upward pressure on prices. More precisely, the indirect mechanism relies on two links: (1) the creation of a monetary-induced gap between the expected rate of profit on capital investment and the market rate of interest and (2) an investment response to this gap. The direct and indirect mechanisms provide the two main channels through which the dynamic price adjustment process works.<sup>1</sup>

**The Neutrality Postulate** A third proposition states that, except for transitional adjustment periods, monetary changes exert no influence on real economic variables, e.g., total output, employment, and the product-mix. These variables, it is argued, are determined by basic non-monetary conditions such as tastes, technology, resource endowments, and rates of technical substitution between factor resources. As the quantity of money in no way alters these fundamental conditions, it follows that monetary changes

<sup>1</sup> Two points of clarification should be made here. First, one does not necessarily have to be a quantity theorist to accept the validity of the monetary transmission mechanisms. In fact, the indirect mechanism today is frequently associated with non-quantity approaches to monetary theory. Second, modern quantity theorists sometimes argue for the direct money-spending mechanism merely as an empirical proxy for a complicated portfolio adjustment process in which specific interest rate effects cannot be captured statistically.

are neutral in their long-run effects on real variables. In brief, money is thought to be merely a veil, obscuring but not affecting the operation of real economic forces.

Note, however, that the neutrality postulate, like the proportionality postulate, refers only to long-run equilibrium. During the short-run transition to equilibrium, monetary changes very definitely can have non-neutral effects on real variables. For example, during the transition period there may be wealth-distribution effects stemming from the failure of some cashholders to get their pro rata share of additional money and from the impact of unanticipated price-level changes on the real value of fixed-dollar financial claims. These distribution effects will alter the composition of demand and thus the structure of production. Moreover, some commodity and factor prices may adjust more swiftly than others thereby altering relative prices (market exchange ratios) and thus relative quantities of real variables. The quantity theory does not deny that money changes may influence resource allocation in the transition period. What it does claim, however, is that these non-neutral effects are temporary and that they will vanish in long-run equilibrium when the economy has adjusted fully to the monetary change.

**Monetary Theory of the Price Level** The neutrality postulate states that changes in the quantity of money affect only the price level. As stated, however, this proposition is not sufficient to rule out the possibility that non-monetary variables may also be important determinants of  $P$ . An additional condition must be invoked. Accordingly, a fourth postulate states that the price level itself tends to be influenced predominantly by changes in the quantity of money. The implication is that price level instability stems principally from monetary rather than non-monetary disturbances. Thus, inflation and deflation are largely attributed to the erratic behavior of the money stock rather than to non-monetary causes originating in the real (commodity) sector of the economy.

It should be noted that the fourth postulate refers to the general price level and not to relative prices, i.e., relationships among the prices of individual commodities (market exchange ratios). Quantity theorists readily admit that non-monetary influences—e.g., technological progress and productivity change; crop failures, embargoes, and other disruptions in the supplies of food and raw materials; monopoly power; excise taxes and the like—can directly affect relative prices. But they argue that such non-monetary-induced changes in the prices of some commodities are often likely to be balanced by opposite changes in

the prices of others, leaving the average price level unchanged. They hold that it is usually monetary shocks, not real-sector disturbances, that exert the dominant effect on the general level of prices.

**Exogeneity of the Nominal Stock of Money** A fifth condition required by the quantity theory is that the nominal stock of money be non-demand determined. This requirement is a corollary of the proposition that nominal  $M$  is the independent causal factor governing  $P$ . For if the quantity of money is not an independent variable, but instead responds passively to prior shifts in the demand for it, then quantity theorists could not claim that it played the active initiating role in the determination of the price level.

It should be emphasized that the exogeneity postulate refers to the *nominal* rather than the *real* stock of money. The distinction between the two stocks is crucial. Unlike the nominal stock, the real stock is treated by the quantity theory as an endogenous variable determined by the public's demand for real balances. As previously discussed, the public, via the impact of its spending on the price level, can make the real value (purchasing power) of any given nominal stock of money equal to the desired quantity of real cash balances. In brief, the real money stock is seen as a dependent variable determined by the public's decisions to acquire or get rid of cash.

Such is not the interpretation given to the nominal stock, however. Quantity theorists long have argued that, in fact, the nominal stock of money is largely determined by factors independent of those determining the demand for it. Traditionally, the quantity theory has treated the nominal money stock as a largely exogenous variable. In the days of the gold specie standard, a nation's money stock was regarded as mainly predetermined by the past and current production of gold and by the state of the external accounts (balance of payments). Later, when paper money had replaced gold, the stock of money was regarded as exogenously determined by the independent central bank via its control over a narrowly-defined base of so-called high-powered money consisting chiefly of the central bank's own liabilities. This interpretation of the central bank as the exogenous controller of the money stock, it should be pointed out, assumes the existence of stable links between the base of high-powered money created by the central bank, and the deposit and banknote money generated by the commercial banking system. These stable links are necessary if the total money supply is to behave exactly as its exogenously determined component, the monetary base. Generally, quantity theorists have argued that these stable links exist.

Quantity theorists also have employed the notion of stable linkages to minimize the problems that money substitutes may pose for monetary regulation and control. The quantity theory has never denied that near-moneys may influence spending and prices just as money does. What the theory has denied, however, is that the volume of money substitutes can expand or contract independently of the volume of money and thus act as an autonomous influence on the price level. Instead, money and money substitutes are thought to be related via a stable link so that variations in the former will be accompanied by roughly proportional variations in the latter.

### DEVELOPMENT OF THE QUANTITY THEORY UP TO THE NINETEENTH CENTURY

The main outlines of the quantity theory began to take shape as early as the mid-16th century when Jean Bodin first stated his monetary theory of the price inflation then occurring in Western Europe. Later writers lent precision to Bodin's hypothesis by postulating that the value or purchasing power of money varies in exact proportion to the quantity in circulation so that a doubling of  $M$  will double  $P$  and halve the value of the monetary unit. At first the proportionality postulate was treated as an identity. As originally stated by John Locke in 1691, the postulate asserted that  $P$  is *always* proportional to  $M$ . In 1752 David Hume introduced the notion of causation by stating that variations in  $M$  will *cause* proportionate changes in  $P$ . By the time it reached the Classical economists in the early 19th century, the proportionality postulate was understood as a proposition of comparative static analysis, valid only for the comparison of states of old and new monetary equilibrium after the economic system had fully adjusted to a change in the money stock. It was recognized that proportionality between  $M$  and  $P$  temporarily would be disturbed during the transition period between successive monetary equilibria.

Richard Cantillon and David Hume, both writing in the 18th century, were the first to apply to the quantity theory the two crucial distinctions: (1) between economic statics and dynamics, i.e., between long-run stationary equilibrium and short-run movements toward equilibrium, and (2) between the long-run neutrality and the short-run non-neutrality of money. In what were perhaps the earliest examples of dynamic process analysis, these writers described the sequence of steps by which the impact of a monetary change spreads from one sector of the economy to another, altering relative prices and quantities in the process. Cantillon and Hume pointed out that

adjustment would continue until all prices had changed in equal proportion to the money stock and all quantities had returned to their pre-existing levels. Especially vivid was the Cantillon-Hume account of the short-run non-neutrality of money. Cantillon pointed out that the dynamic adjustment path would be influenced by the way new money was injected into the system. Specifically, he stated that most monetary injections would involve non-neutral distribution effects. He argued that, generally, new money will not be distributed among individuals in proportion to their pre-existing share of money holdings. Some will receive more, and others less, than their proportionate share. The former group will benefit at the expense of the latter and therefore, via their money outlays, will play a greater role in determining the composition of output. In short, Cantillon demonstrated how initial distribution effects temporarily could alter the pattern of expenditures and thus the structure of production and the allocation of resources.

David Hume described how different degrees of money illusion among income recipients, coupled with time delays in the adjustment process, could cause costs to lag behind prices, thus creating profits and stimulating the formation of optimistic profit expectations. Hume believed that both actual profits and optimistic expectations would spur business expansion and employment during the transition period. These non-neutral effects were expected to vanish in the long-run, however.

To the Cantillon-Hume list of temporary non-neutral effects, 19th century economists added others, most of which stemmed from the fixity of certain types of contractual payments and from the failure of all factor-resource prices to adjust with equal swiftness. These additional non-neutral effects included: (1) the lag of money wages behind prices which temporarily reduces real wages, thereby encouraging increased demand for labor; (2) the stimulus to output occasioned by inflation-induced reductions in real debt burdens which shift real income from unproductive creditor-rentiers to productive debtor-entrepreneurs; (3) so-called "forced-saving" effects, i.e., changes in the fraction of the economy's resources diverted from consumption into capital formation owing to price-induced redistributions of income among socio-economic classes having different propensities to save and invest; and (4) the stimulus to investment spending imparted by a temporary reduction in the loan rate of interest below the profit rate on real capital.

While acknowledging the existence of these non-neutral effects, however, classical quantity theorists

frequently tended to minimize their importance. This de-emphasis of transition effects is what distinguished the classical version of the quantity theory from the earlier Cantillon-Hume version. Whereas the latter tended to stress dynamic disequilibrium periods in which money matters much, classical analysts focused on long-run equilibrium in which money is just a veil. Whereas Cantillon and Hume thought that transition periods would be protracted, classical analysts saw them as evanescent. Whereas the Cantillon-Hume analysis stressed the output and employment effects of inflation, classical analysis virtually ignored, or treated as insignificant, those real effects. The prevailing view, the position of the most influential of the classical economists, especially David Ricardo, was that these disequilibrium effects were ephemeral and unimportant, mere qualifications to the long-run equilibrium analysis. This opinion may have been conditioned by Ricardo's penchant for abstract, comparative-static theorizing. Or perhaps it sprang from his desire for an uncomplicated and convincing theory to support his charge that inflation in Britain was solely the result of the Bank of England's irresponsible overissue of currency. Such a theory would be more effective if it isolated price-level effects and abstracted from real effects. Most likely, Ricardo and other classical economists avoided discussions of any beneficial output and employment effects of monetary injections in fear of providing crude inflationists with arguments to support their case. Whatever the reason, non-neutral transition effects were slighted.

Finally, an advance in understanding of the monetary transmission mechanism occurred. This progress accompanied the historical evolution from a predominantly full-bodied money to a mixed metal-paper money that occurred in the 18th century. Written in the era of full-bodied money, the Cantillon-Hume account of the adjustment process had relied solely on the direct mechanism to raise prices. In the Cantillon-Hume analysis, an arbitrary influx of gold coin induces an increase in the rate of spending until all incomes and prices had risen in proportion to the monetary injection. The direct mechanism, however, no longer sufficed as an explanation of the adjustment process after gold coin had given way to bank notes in the 19th century. The main shortcoming of the direct mechanism was that it failed to explain how bank notes and other forms of paper money are injected into the system. In his 1802 classic, *The Paper Credit of Great Britain*, Henry Thornton provided the first exposition of the indirect mechanism. Pointing out that new money created by banks enters the financial markets initially via an

expansion of bank loans, Thornton described how the increased supply of loanable funds temporarily reduces the loan rate of interest below the profit rate (expected yield) on new capital projects. This disparity between profit and loan rates stimulates additional investment spending, thereby exerting upward pressure on product prices, including the price of investment goods. With investment goods becoming increasingly expensive, however, businessmen require more and more loans to finance their purchases. The demand for loans therefore increases, bidding up the loan rate of interest in the process. Equilibrium is reestablished when rising loan demand eventually overtakes the initially expanded supply and the money rate of interest rises back into equality with the profit rate. Nineteenth century quantity theorists incorporated both the Cantillon-Hume direct mechanism and the Thornton indirect mechanism in their explanation of the linkages between M and P.

#### ROLE OF THE QUANTITY THEORY IN CLASSICAL POLICY DEBATES

The first half of the 19th century, an era in which the doctrines of the British classical school dominated economic thought, saw the emergence of a concentrated and systematic application of the quantity theory to policy problems. Having been quickly absorbed into the mainstream of classical analysis, the quantity theory became the standard conceptual framework for the analysis of monetary problems and for the formulation of practical policy recommendations. The central monetary problems in England at that time related to the maintenance of external equilibrium and the restoration and preservation of the gold standard. Consequently, the quantity theory tended to be directed toward the analysis of international price levels, gold drains, exchange rate fluctuations, trade balance deficits, and related problems.

**The Quantity Theory and the Price-Specie-Flow Mechanism** It was only natural that the quantity theory was applied to these problems of international finance. After all, the theory had long played a strategic role in the classical theory of international trade. The quantity theory was the key ingredient in the classical explanation of the operation of the price-specie-flow mechanism, i.e., the automatic self-regulating adjustment mechanism that insures the restoration and preservation of balance of payments equilibrium and that governs the international distribution of the precious metals. One of the earliest rigorous explanations of the specie-flow mechanism was provided by David Hume. In one of the more celebrated pas-

sages in British economic literature, Hume started out by assuming a five-fold overnight increase in the domestic money supply. Proceeding to trace the consequences, he argued that wages and prices would rise in proportion to the monetary change, thereby making British goods more expensive than foreign goods, and thus causing imports to rise, exports to fall, and gold to flow out. The external gold drain, in turn, would tend to moderate prices in Britain and raise them elsewhere. Hume held that the trade-balance deficit and the specie outflow would continue until the purchasing power of gold was the same everywhere, imports and exports were in balance, and the terms of trade were identical to those that would reign under a purely barter regime.

It is readily apparent that Hume's explanation embodies most of the key elements of the quantity theory. The proportionality postulate is stated explicitly. The most prominent element, however, is the interpretation of money as the active causal variable—disturbing initial equilibrium, driving up prices, generating a trade imbalance, forcing an efflux of specie, and eventually restoring equilibrium. The short-run non-neutrality of money emerges in the form of the alteration of the terms of trade or relative prices of exports and imports. And the long-run neutrality of money is manifested in the restoration of the pre-existing barter ratios. These same quantity theory elements comprised the analytical framework within which classical economists discussed the events and policy issues surrounding the leading monetary controversies of the day.

**The Bullionist Controversy** The two great monetary debates of the classical era were (1) the Bullionist controversy that took place in the first two decades of the 19th century during and immediately after the Napoleonic Wars and (2) the Currency School-Banking School controversy during the middle decades of the century. The Bullionist controversy was provoked by events following a major policy shift in 1797. In that year, under the stress of the Napoleonic Wars, Britain left the gold standard for an inconvertible paper standard. A series of gold drains, coinciding with heavy military outlays abroad, extravagant government borrowing, and extraordinary wheat imports, had virtually exhausted the Bank of England's gold holdings. The depletion of the Bank's gold reserve thus forced the suspension of specie payments. The Bank was released from its obligation to exchange gold for currency at the fixed mint price, i.e., bank notes were no longer automatically convert-

ible into gold. The suspension of specie payments was followed by a rise in the price of bullion, foreign exchange, and commodities in terms of paper currency. A debate then arose centering on the following issues: Was the pound depreciated? Was there inflation in Britain and if so, what was its source?

The Bullionists, led by David Ricardo, argued that currency depreciation and inflation did exist, that the overissue of bank notes by the Bank of England was its cause, and that the premium quoted on bullion (the difference between the market and the old mint price of gold in terms of paper money) was the proof. Price indexes not being in use then, the Bullionists used the gold premium as we use price indexes today to measure the extent of inflation.

The Bullionists arrived at their conclusions via the following route: the quantity of money determines domestic prices; domestic prices affect the exchange rate; and the exchange rate between inconvertible paper and gold standard currencies determines the premium on gold. It follows, therefore, that the depreciation of the exchange rate below gold parity (i.e., below the ratio of the respective old mint prices of gold in each country) and the existence of the premium on bullion both constituted evidence that prices were higher and the quantity of money greater in Britain than would have been the case had the country still been on the gold standard.

In short, the depreciation of both the internal and the external value of the paper pound was attributed solely to the redundancy of money, and the Bank of England was reproached for having taken advantage of the suspension to expand its note issues recklessly. Thus, like present day monetarists, Bullionists located the source of inflation in the central bank. But the Bullionists went even further, charging that the Bank was also responsible for the external specie drains that led to the restriction of cash payments. Bullionists claimed that the redundancy of notes, by forcing up domestic prices relative to foreign prices, had caused the trade balance to become adverse, thus forcing gold to leave the country. Here is the quantity theory view of money as the active disturber of economic equilibrium.

The same quantity theory reasoning underlay the Bullionists' policy prescription for restoring convertibility. Bullionists held that the sole prerequisite for the restoration of specie payments at the old mint price was the contraction of the note issue. The assumption was that the reduction in the money supply would lower internal prices, remove the trade deficit, bring the exchanges back to par, and eliminate the premium on bullion. With sufficient reduction of the note issue, convertibility could be restored without

fear that an external drain of gold would again deplete the country's bullion reserve.

**Control of the Money Supply** In the main, the theory employed by the Bullionists in locating the source of inflation was the same quantity theory that they had inherited from their pre-classical predecessors. It would be wrong, however, to assume that the Bullionists did not add anything to the theory. Their specific contribution related to the question of the control of the money supply. They were the first to develop the idea that the stock of money, or at least the currency component, could be effectively regulated via the control of a narrowly defined monetary base. This point was first brought out in their treatment of the relation between the volume of Bank of England notes and the note issues of the country banks. The money supply at that time consisted of gold coin and the note liabilities of both the Bank of England and country banks. The link between the entire money supply and the Bank of England's note component might have appeared tenuous because of the possibility of the country bank note component expanding and contracting independently of Bank of England notes. But the Bullionists denied this possibility on two grounds. First, the country banks tended to keep as a reserve Bank of England notes equal to a relatively constant percentage of their own note liabilities. Second, any overissue of country bank notes (and consequent rise in local prices relative to London prices) would drain Bank note reserves from the countryside to London via a regional balance of payments or specie-flow mechanism, thereby forcing the country banks to contract their note issues. For these reasons, asserted the Bullionists, country bank notes would be passively tied to Bank of England notes by a virtually rigid link and could expand or contract only if the Bank's own issues did. This is the origin of the quantity theory view that control of a narrowly-defined base of "high-powered money" implies virtual control of the money supply.

**The Currency-Banking Debate** The second great 19th century debate in which the quantity theory played a leading role was the Currency-Banking controversy over the question of the regulation of the bank note issue. This debate took place in the 30-year period following Britain's return to the gold standard in 1821. The main policy objectives of this period included maintenance of fixed exchange rates and the automatic gold convertibility of the pound. Members of the Currency School, applying the precepts of their Bullionist forebears, held that such preservation of the gold standard could be secured

only through rigid adherence to the "Currency Principle" of making the existing mixed gold-paper currency behave exactly as would a wholly metallic currency, i.e., by requiring banknotes to expand and contract one for one with variations in gold reserves. Given the desirability of making paper money behave exactly like a metallic one, however, by what means or device was this result to be achieved? By the mere requirement of gold convertibility alone? Or by the imposition of even stricter rules and regulations on the note issue? These questions constituted one of the central issues of the controversy.

**Safeguards to Note Overissue: Convertibility vs. Regulation** The Bullionists had argued earlier that convertibility as such would be sufficient to insure that banknotes would respond automatically to gold flows in conformity with the principle that the mixed currency should behave like a metallic one. Convertibility alone, Bullionists thought, would be an adequate safeguard against overissue. If too many notes were issued, they reasoned, then according to the quantity theory the value of the notes would fall and the foreign exchange rate would depreciate. People would then convert notes into gold for export, and the consequent loss in specie reserves would force the Bank to contract its note issues.

Members of the Currency School, however, regarded convertibility as an inadequate check to overissue. They feared that even a legally convertible currency would be issued to excess with the following unfortunate consequences: rising domestic prices relative to foreign prices; unfavorable balance of payments; weakened foreign exchange; gold outflow; depletion of gold reserves; and ultimately, suspension of convertibility. The rate of reserve depletion would be accelerated, they noted, if the external gold drain coincided with an internal drain as domestic residents, alarmed by the possibility of suspension, sought to convert paper currency into gold.

**Time Lags and Destabilizing Policy Responses** The apprehensions of the Currency School stemmed from its belief that the past actions of the Bank of England had been perverse and destabilizing. This destabilization argument stressed the adverse effects of time lags on the Bank's policy response to gold outflows and to exchange rate movements. Specifically, the Currency School argued that long lags existed between changes in the volume of notes outstanding and consequent changes in prices and the exchange rate. Owing to these lags, the exchange rate would be slow in registering the effect of a note overissue and in signaling the need for a corrective contraction. Guided by the exchange rate indicator,

the Bank would continue to expand its specie-displacing note issues long after the appropriate time for contraction.

In sum, the Currency School contended that long time-lags affected the Bank's policy response to gold drains. Because of these delays, the Bank's reactions to external drains often came too late to protect the specie reserve and served instead to weaken public confidence in the Bank's ability to maintain convertibility. Moreover, when the Bank finally *did* apply restrictive policies to stem the gold losses, these policy actions tended to coincide with and to exacerbate the financial panics and liquidity crises that inevitably seemed to follow periods of currency and credit excess. In short, the Currency School alleged that the Bank's policy actions had accentuated, rather than alleviated, economic disturbances. These Currency School arguments foreshadowed by more than 100 years Milton Friedman's doctrine that the prevalence of long lags in the response of spending and prices to changes in the money supply, and, to a lesser extent, in the policymakers' response to changes in the economy, tend to render discretionary stabilization efforts destabilizing. What was needed to prevent the recurrence of gold drains, exchange depreciation, and domestic liquidity crisis, the Currency School thought, was convertibility *plus* strict regulation of the volume of Bank notes.

**Policy Prescriptions of the Currency School** The Currency School was successful in exacting its ideas into legislation. The famous Bank Charter Act of 1844 embodied the prescription that, except for a small fixed amount of notes that the Bank could issue against government securities, new notes could be emitted only if the Bank had received an equivalent amount of gold. In modern terminology, the Charter Act established a marginal gold reserve requirement of 100 percent behind note issues. With notes rigidly tied to gold in this fashion, external gold drains would be accompanied by reduction of a like amount of notes domestically.

The quantity theory clearly underlay the Currency School's prescription for stabilizing prices, securing convertibility, and preserving the gold standard by tying the note issue to gold. For this prescription was based on the postulate that money stock changes cause price level changes. The Currency School held that the channel of influence ran from domestic note overissue to rising prices to a weakened trade balance and deterioration of the foreign exchanges and, ultimately, to gold outflows. Similarly, domestic price rises would be reversed and the foreign exchanges strengthened by reducing the note issue. By tying

notes to gold with a 100 percent reserve requirement, the money stock would be regulated and, consequently, the stability of the external value of the pound would be achieved automatically.

**Money Substitutes and the Effectiveness of Monetary Control** The Currency School also contributed to the quantity theory doctrine that money substitutes cannot impair the effectiveness of monetary regulation. This proposition is based on two underlying presumptions: (1) that money, the specific control instrument, can be clearly identified and unambiguously distinguished from money substitutes and (2) that money and near-money are related via a stable link so that variations in the former will be accompanied by predictable variations in the latter. These points were brought out in the Currency School's treatment of bank notes versus other forms of circulating media. At a time when bills of exchange and bank deposits were being employed increasingly as instruments of exchange, Currency School advocates concentrated solely on notes. They insisted that money be defined to include only coins and notes and that monetary regulation be confined to notes. They felt justified in excluding near-money—bills of exchange and bank deposits—from their policy analysis. They thought that the entire superstructure of money substitutes could be regulated effectively by control of the money (bank note) base. In particular, they thought that the limitation on note issues constituted an ultimate constraint on the creation of deposits. Hence, rigid control of the former implied equally rigid control of the latter. Thus, if notes could be controlled, there would be no need for explicit control of deposits. They defended their sharp distinction between money (coin and notes) and near-money (deposits and bills) on two grounds. First, the low circulation velocity of near-moneys rendered those instruments quantitatively insignificant relative to notes as exchange media. Second, in times of financial crises near-moneys were poor substitutes for money strictly speaking, because only the latter would be accepted in final payment. Here, in the Currency School's analysis, is the origin of quantity theorists' tendency (1) to make a sharp delineation between money and other liquid assets and (2) to deny that near-moneys can frustrate the effects of changes in the money supply.

#### ANTI-QUANTITY THEORY VIEWS

**A Catalogue of Criticisms** There has long been a body of doctrine opposed to the quantity theory. At one time or another each of the following criti-



cisms have been leveled against the theory. These criticisms are neither mutually exclusive nor are they always consistent: There is some overlapping and some conflict.

(1) Modern Keynesians argue that the quantity theory is invalid because it assumes an automatic tendency to full employment. If resource unemployment and excess capacity exist, a monetary expansion, if effective, may produce a rise in output rather than a rise in prices. More generally, money may be more than just a veil. Monetary changes may have a *permanent* effect on output, interest rates, and other real variables, contrary to the neutrality postulate of the quantity theory.

(2) Post-Keynesian economists also contend that the quantity theory erroneously assumes the stability of velocity and its counterpart, the demand for money. In fact, velocity is a volatile, unpredictable variable, influenced by expectations, uncertainty, and by changes in the volume of money substitutes. The erratic behavior of velocity makes it impossible to predict the effect that a given monetary change will have on prices. Changes in velocity may offset (negate) or accentuate the price-level impact of a monetary change.

(3) Nineteenth century proponents of the so-called *real bills* doctrine argued that, contrary to the quantity theory, the money supply is an endogenous variable that responds passively to shifts in the demand for it. One implication is that monetary changes cannot influence prices. Being demand-determined, the stock of money cannot exceed or fall short of the quantity of money demanded. And with the quantity of money supplied always identical to the quantity demanded, no situation of excess supply or redundancy of money can ever develop to stimulate spending and force up prices. In short, there is no transmission mechanism running from money to prices.

(4) In fact, claimed real bill advocates, the channel of influence runs in the opposite direction. Causation flows from prices and income to money, rather than vice versa. Income and prices determine the demand for money, which, in turn, determines the money supply. And since the money supply is the result and not the cause of variations in income, prices and spending, it follows that monetary changes cannot be the source of inflation, deflation, and other economic disturbances. Hence, the quantity theorists' monetary interpretation of inflation, balance of payments disequilibrium, and business cycles must be wrong. Analysts should instead seek for the sources of economic disruptions in real (non-monetary) causes.

(5) A host of critics, both modern and old, have maintained that, contrary to the quantity theory, a monetary injection cannot always be relied upon to stimulate spending and increase prices. A monetary expansion may be ineffective for at least three reasons. First, the new money may simply be absorbed into idle hoards. Second, spending may be interest-insensitive, i.e., unresponsive to changes in interest rates induced by the monetary expansion. Third, as previously mentioned, the money stock may be demand-determined, in which case there can be no excess supply of money to spill over into the commodity market in the form of an excess demand for goods.

Many of these criticisms originated in the contra-quantity theory doctrines of the 19th century adversaries of the Bullionists and the Currency School.

**Antibullionist Opposition to the Quantity Theory**  
Opposition to the quantity theory emerged early in

the Bullionist debate in the form of the Antibullionists' critique of the Bullionists' policy analysis. At least two contra-quantity theory arguments can be identified in the Antibullionist position. First is the rejection of a monetary for a non-monetary explanation of economic disturbance. In opposition to the Bullionists' contention that both the gold premium and the depreciation of the paper pound were attributable to the overissue of currency, the Antibullionists maintained that the rise in the prices of bullion and foreign exchange were due to an unfavorable balance-of-payments stemming from non-monetary causes, notably domestic crop failures and heavy military outlays abroad. Moreover, Antibullionists denied that excessive money creation was the cause of the gold outflow and suspension of convertibility. Similarly, they doubted that mere contraction of the note issue would be sufficient to permit resumption. They argued that reduction of imports and curtailment of war-related foreign expenditures were the essential prerequisites for the restoration of the gold standard. This argument is the essence of the anti-quantity theory view that economic disturbances stem from non-monetary causes and require non-monetary cures.

Second, Antibullionists employed the real bills doctrine to assert the impossibility of an excess supply of money ever developing to exert upward pressures on prices. The real bills doctrine states that just the right amount of money and credit will be created if bank loans are made only for productive (nonspeculative) purposes. Defending the Bank of England against the Bullionists' charge of note over-issue, Antibullionists argued that excessive issues were impossible as long as the Bank's note liabilities were based on sound commercial paper, i.e., were issued only to finance genuine production and trade. The real bills criterion, Antibullionists contended, would insure that the volume of currency would adapt itself automatically to the needs of trade. Bank notes issued to finance the production of goods would be extinguished when the goods were marketed and the real bills were retired (loans were repaid) with the sales proceeds. Since money creation would be limited to the expansion of real output, no inflation could occur. Here is the origin of the contra-quantity theory notion that the stock of money is solely demand-determined and therefore can have no independent influence on spending and prices.

**Anti-Quantity Theory Views of the Banking School** The main attack on the quantity theory, however, was launched by the Banking School in its debate with the Currency School. Led by Thomas Tooke, John Fullarton, and James Wilson, Banking

School analysts challenged the validity of virtually all of the propositions of the quantity theory. They denied that monetary expansion or contraction would affect prices. They argued that changes in the supply of money and credit could not be expected to influence spending and prices for two reasons. First, new money may simply be absorbed into idle balances without entering the spending stream. Second, the supply of money is determined by the needs of trade and thus can never exceed demand.

The first point was brought out in the discussion of gold hoards. The Banking School alleged that great accumulations of idle money existed in the form of hoards of precious metals. These hoards supposedly were held mainly by the banks as excess bullion reserves. The full impact of gold flows, it was argued, would be absorbed entirely by those hoards without affecting the amount of currency in circulation. Imports of monetary gold would augment the hoards without causing an increase in the circulating media or inducing a rise in spending. Similarly, an outflow of gold would be withdrawn from the bullion hoards, but would have no effect on the monetary circulation or the volume of expenditure.

The second point was brought out in the Banking School's discussion of the real bills doctrine and the *law of reflux*. Like its Antibullionist predecessors, the Banking School contended that currency over-issue was impossible as long as banks restricted their loans to self-liquidating commercial or agricultural paper. But the Banking School went further than the Antibullionists, arguing that even if the real bills criterion were violated, the law of reflux would operate to prevent overissue. If notes were emitted in excess of legitimate working capital needs, the public would not wish to hold the excess notes and would deposit them, use them to repay bank loans, or redeem them for coin. In any case, the excess notes would be returned immediately to the banks. In brief, the real bills criterion together with the reflux mechanism would provide a sufficient check to over-issue. Notice how the Banking School, in rejecting the possibility of an excess supply of money and credit, also denied the validity of the monetary transmission mechanism propounded by the quantity theory. According to the latter, an excess supply of money is what induces the excess demand for goods that bids up prices, i.e., following a monetary injection, people try to get rid of undesired additional money holdings by spending them. This adjustment mechanism, however, was implicitly denied by the Banking School's insistence that the supply of money is always identically equal to the demand for it.

In its opposition to the quantity theory, the Bank-

ing School developed its own non-monetary theory of the price level. Thomas Tooke stated explicitly that the general level of prices was determined by *incomes* (wages, rents, profits, etc.), and not by the quantity of money. Tooke's argument was that factor incomes, rather than money, are the sources of expenditures that act on prices. This is an early version of the income-expenditure approach to monetary theory, an approach that formulates its analysis in terms of the determinants of aggregate demand rather than in terms of the quantity of money or the velocity of circulation. The income-expenditure approach was later developed by Keynes, and continues to be a characteristic feature of Keynesian macroeconomic models.

Tooke did not explain how these price-determining factor incomes themselves were determined but left the question of their origin open to a variety of possible interpretations. His theory of price inflation is therefore also suggestive of recent wage-cost-push and structural theories that (1) link inflation to some arbitrary non-monetary element in the institutional environment, e.g., autonomous increases in wage incomes, production bottlenecks, particular supply inelasticities, institutional price rigidities, etc., and (2) stress the inflationary role of the competitive struggle for relative shares in the national income.

Factor incomes were not the only price-influencing forces discussed by the Banking School. Changes in profit anticipations also were mentioned frequently. What was stressed, however, was the non-monetary nature of these expectational influences. This emphasis reflects the contra-quantity theory tendency to attribute price level movements to non-monetary forces rather than to changes in the money supply.

The Banking School also disputed the quantity theory view of money as an exogenous or independent variable. Banking School writers argued that the stock of money and credit is a passive, endogenous, demand-determined variable—the effect, not the cause, of price changes. Contrary to the Currency School's contention that the channel of influence runs from money to prices, the Banking School argued that the channel of causation runs in the opposite direction. That is, when prices, total money income, and aggregate demand are increasing, the demand for loans would rise and the banking system would accommodate the increased loan demand by supplying additional credit and circulating media. In the determination of the volume of currency in existence, the non-bank public (borrowers) played the active role and banks (issuers of money) the passive or accommodating role. Implicit in the Banking School's view of passive money are three anti-quant-

tity theory propositions: (1) changes in economic activity precede and cause changes in the money supply (the so-called reverse-causation argument), (2) the supply of circulating media is not independent of the demand for it, and (3) the central bank does not actively control the money supply but instead accommodates or responds to prior changes in the demand for money.

Concerning the problem of money and money substitutes, the Banking School disputed the quantity theory view that control of the former implied control of the latter. Contrary to the Currency School's stress on a narrowly-defined money supply, the Banking School tended to emphasize the overall structure of credit. The Banking School criticized the Currency School's attempts to draw a hard and fast line between money and near-money. The Banking School argued that the ready availability of bank deposits, bills of exchange, and other forms of credit instruments that could circulate in lieu of money would defeat the Currency School's efforts to control the entire credit superstructure via control of the banknote base. The Banking School thought that the volume of credit that could be erected on a given monetary base was large, variable, and unpredictable. The total volume of credit, it was argued, is independent of, as well as quantitatively more significant than, the money stock. Here is an early example of two more anti-quantity theory notions, i.e., (1) the difficulty of making a watertight distinction between money and near-moneys, and (2) the ineffectiveness of policy attempts to stabilize prices via control of the stock of money in a financial system that can produce an endless array of money substitutes.

The contra-quantity theory views of the Banking School strongly influenced its position on at least three important policy questions. First, on the question of free versus regulated banking, the Banking School advocated more free trade and less regulation in banking than did the Currency School. The Banking School thought that the quantity of money and credit would best govern itself automatically through the force of people's self-interest. Thus, if the supply of money is determined by the needs of trade and is automatically regulated by the reflux mechanism, there was no need for intervention in the form of government legislation such as that proposed by the Currency School. Second, on the question of rules versus discretion in the control of the money supply, the Banking School generally was in favor of discretionary judgment of bankers as opposed to rules of government. The Currency School had advocated that discretionary policy be replaced by a fixed rule, i.e., the 100 percent marginal reserve requirement for

banknote issues. But the Banking School held that banks should not be constrained by a rigid rule, because the optimum quantity of money would be forthcoming automatically if the banks themselves regulated their note and deposit liabilities by responding to the needs of trade. Third, on the question of the rationale of monetary policy, the Banking School regarded attempts to regulate prices via monetary control as both futile and pointless. In the first place, the money supply (especially the note component) is an endogenous variable not subject to exogenous control. And even if the narrow money supply could be controlled, the total paper circulation (total credit), a more comprehensive magnitude that is interchangeable with money, cannot be so controlled. Finally, the Banking School argued that to propose regulation of the price level via control of money and credit is to put the cart before the horse. For it is prices that determine the quantity of money and credit, and not vice versa.

#### THE NEO-CLASSICAL REFORMULATION

Despite the Banking School's criticisms, the quantity theory emerged from the mid-19th century Currency-Banking debate to command widespread acceptance. Moreover, in academic circles at least, it continued to reign as the dominant monetary theory until the 1930's. Several factors may have contributed to the success of the theory. For one thing, the Currency School's policy recommendations of fixed exchanges, maintenance of the gold standard, currency convertibility, and strict control of banknotes became part of British monetary orthodoxy in the second half of the nineteenth century. Since the quantity theory had provided the theoretical foundation for these policy prescriptions, it was only natural that it also was elevated to the rank of established orthodoxy. Then, too, there may have been some decline in the prestige of the opposing real bills doctrine. Long before the end of the century quantity theorists had exposed the fallacies of the real bills criterion as an automatic regulator of the money supply. Quantity theorists had demonstrated that as long as the loan rate of interest is below the expected yield on new capital projects, the demand for loans would be insatiable. In such a case the real bills criterion would provide no effective limit to the quantity of money in existence. Probably the most important contributing factor, however, was the rigorous mathematical restatement of the quantity theory provided by neo-classical economists around the turn of the century. Representing a substantial refinement, systematization, and extension of the earlier Classical

analysis, the neo-classical reformulation added substantially to the intellectual appeal and scientific prestige of the theory.

**Neo-Classical Contributions** The neo-classical reformulation of the quantity theory consisted of at least three separate contributions. First, there was the mathematical framework that neo-classical economists employed to expound and empirically test the key propositions of the theory. This mathematical framework took two alternative forms, namely (1) Irving Fisher's famous *equation of exchange*,  $MV = PT$ , where  $M$  is the stock of money,  $V$  is velocity of circulation,  $P$  is the price level, and  $T$  is the physical volume of market transactions; and (2) the celebrated *Cambridge cash balance equation*,  $M = kPy$ , where  $M$  is the stock of money in circulation,  $k$  is the desired cash balance ratio, i.e., the ratio of the nominal money supply to nominal income,  $P$  is the price level of the national product, and  $y$  is real national income or the national product valued at constant prices. Using these equations, neo-classical analysts were able to spell out precisely the conditions that must hold if the proportionality postulate is to be valid.

As explained by the neo-classical quantity theorists, these conditions included constancy of the velocity of money and of real output. Neo-classical economists held that velocity was a near-constant determined by individuals' cash-holding decisions in conjunction with technological and institutional factors associated with the aggregate payments mechanism. More specifically, it was argued that individuals would try to keep non-interest-bearing transactions balances to the minimum necessary to finance day to day purchases and to provide a reserve for contingencies. The minimum balances that individuals would need to hold, and by implication, the rate of circulation of money, would depend on such factors as the state of development of the banking system, frequency of receipts and disbursements, length of the payment period, degree of synchronization of cash inflows and outflows, rapidity of transportation and communication, etc. Since these factors were thought to be subject to only gradual, evolutionary change, both velocity and the Cambridge  $k$ , it was argued, could be treated as virtual constants in the neo-classical quantity equations. Output and transactions, too, were regarded as constants determined by full-capacity utilization of available resources and technology.

The policy implications of the neo-classical formulation were clear: monetary policy could be expected to exert a powerful, predictable influence on

prices. With velocity,  $V$ , and transactions,  $T$ , both regarded as constants, Fisher's equation of exchange could be expressed in a form,  $P = (V/T)M = (\text{constant})M$ , showing a constant proportional relationship between average prices and the money stock. Embodying the proportionality postulate, this expression implied that a policy-engineered percentage change in the money stock would cause the same percentage change in the price level.

The second neo-classical contribution was the formalization, elaboration, and extension of the Bullionist-Currency School ideas on control of the money supply. Irving Fisher, A. C. Pigou, and other neo-classical analysts demonstrated that monetary control could be achieved in a fractional reserve banking system via control of an exogenously determined stock of high-powered money. They argued that the total stock of money (coin and notes) and bank deposits would be a constant multiple of the monetary base. Underlying their argument were the assumptions that banks desire to hold a fixed proportion of their deposits as reserves and that the public desires to maintain a constant ratio of cash holdings to demand deposits. In short, they argued that the stock of money is governed by three proximate determinants: (1) the high-powered monetary base, (2) the bank's desired reserve-to-deposit ratio, and (3) the public's desired cash-to-deposit ratio, and that the monetary base dominates the latter two determinants.

Finally, neo-classical quantity theorists stressed the short-run non-neutrality of money, a topic that had been relatively neglected in the classical analysis. Neo-classical writers integrated the quantity theory into their analysis of business cycles, showing how changes in the quantity of money were a major cause of booms and slumps and how monetary regulation of the price level was a prerequisite to the stabilization of economic activity.

So influential was the neo-classical formulation of the quantity theory that it continued to serve as the standard macroeconomic model in use up to the 1930's. In that decade, however, it encountered heavy criticism and, discredited, was supplanted by the Keynesian income-expenditure model.

## THE KEYNESIAN-MONETARIST CONTROVERSY

The Bullionist and Currency-Banking controversies represent the leading 19th century examples of the recurrent debate over the quantity theory. The leading example of the debate in the present century, however, is the controversy that has been raging since the mid-1930's between the anti-quantity theory

forces composed of John Maynard Keynes, his immediate followers, and later neo- or post-Keynesians on the one hand, and their pro-quantity theory antagonists, the monetarists, on the other. The debate erupted in 1936 with Keynes, in his classic, *The General Theory of Employment, Interest, and Money*, leveling a barrage of criticism at the quantity theory.

**The Keynesian Attack** Keynes's attack on the quantity theory consisted of five interrelated elements. First, he argued that the quantity theory assumed an automatic tendency of the economy to operate at full capacity, an assumption patently at odds with experience in the depressed 1930's. Only if production and employment are fixed at full capacity, said Keynes, would monetary-induced changes in spending manifest themselves solely in price level changes. But, he added, if the economy were operating at less than full employment, with idle resources to draw from, changes in spending would affect output and employment rather than prices. Thus, in much of his analysis of the economics of depression, Keynes reversed the assumptions of the quantity theory, treating prices as rigid and output as flexible. He rationalized his assumption of price rigidity by arguing that prices are governed by wage costs, and that union bargaining strength and other institutional forces prevent wages from being downwardly flexible even in depressions. Thus his argument reflected the anti-quantity theory view that the price level is determined by autonomously given factor costs rather than by the quantity of money.

Second, Keynes criticized the particular version of the quantity theory expressed in the neo-classical quantity equations on the grounds (1) that it was a tautological identity rather than an empirically refutable hypothesis, and (2) that it erroneously treated the circulation velocity of money as a near-constant. Keynes contended that, in actuality, the velocity variable in Fisher's equation of exchange was extremely unstable and that it might passively adapt to independent changes in the other elements of the equation. Thus, said Keynes, the impact of any change in  $M$  might be absorbed by an offsetting change in  $V$  and therefore would not be transmitted to  $P$ . Likewise, any change in income or the volume of market transactions might be accommodated by a change in velocity without requiring any change in the money supply.

Third, Keynes revived the Banking School conclusion concerning the futility of using monetary policy to regulate economic activity. Keynes held that monetary policy would be an ineffective cure

for unemployment and recession for two reasons. First, monetary injections might be absorbed immediately into idle hoards without lowering interest rates sufficiently to stimulate investment spending. This conclusion was based on Keynes's theory of an absolute preference for liquidity at low interest rate levels, i.e., the case of the so-called liquidity trap. The theory of the liquidity trap stated that under certain circumstances — e.g., a severe depression characterized by an abnormally low rate of interest and by virtually unanimous expectations of capital losses owing to anticipated rises in bond yields and declines in bond prices — idle cash balances become perfect substitutes for bonds in wealthholders' portfolios. That is, when the anticipated capital loss on bonds is large enough to at least offset the low current interest return, there would be no inherent advantage to holding bonds rather than zero-yield cash. Consequently, the quantity of money demanded would become insatiable, i.e., infinitely sensitive to the slightest change in the rate of interest. In this liquidity trap case, only minute reductions in interest rates would be necessary to induce portfolio optimizers to hold virtually any amount of additional cash injected into the system. Increases in the money supply, therefore, would be ineffective in reducing interest rates and thus in stimulating investment spending via the interest rate channel. Here is the reappearance of the Banking School argument that a monetary expansion cannot be counted on to stimulate spending because the new money may simply disappear into idle hoards. Second, Keynes argued that even if monetary injections were successful in lowering market interest rates, those injections still would not stimulate economic activity if investment spending was unresponsive to changes in interest rates. To summarize, Keynes argued that either a liquidity trap or an interest-insensitive investment expenditure schedule could render a monetary expansion ineffective in a depression. In terms of Fisher's equation of exchange,  $MV = PT$ , a rise in  $M$  would be offset by a fall in  $V$ , leaving total spending,  $PT$ , unchanged. With variable velocity absorbing all the impact of money stock changes, none would be transmitted to nominal income. The rigid links between money, spending, prices, and nominal income postulated by the quantity theory would be severed or severely weakened.

Fourth, like Thomas Tooke, Keynes argued that the income-expenditure analysis was superior to the quantity theory as an analytical model. Keynes's model emphasized the determinants of expenditure rather than the quantity of money. Moreover, it stressed a new non-monetary adjustment mechanism

— the so-called income multiplier — rather than the old direct and indirect monetary linkages. Specifically, Keynes argued that there is a multiplier relationship between autonomous expenditure (i.e., non-income-induced expenditures, e.g., government outlays for armaments or public works projects) and total income, such that a dollar change in the former will stimulate a two or three dollar change in the latter. The Keynesian emphasis on the determinants of spending rather than the stock of money was clearly in the tradition of Thomas Tooke. The chief policy implication of the Keynesian income-expenditure analysis was that fiscal policy would have a more powerful impact on income and employment than would monetary policy. Accordingly, Keynesians argued that chief reliance should be placed on government budgetary (tax and expenditure) policy rather than on monetary policy to stabilize the economy.

Finally, Keynes adhered to a non-monetary explanation of the Great Depression, arguing that the downswing had been triggered by a collapse of capital spending stemming from the disappearance of profitable investment opportunities, and that the contraction had been intensified by a collapse of confidence. Here is a restatement of the Anti-bullionist-Banking School notion that economic disturbances arise from exogenous shocks originating in the real sector rather than from erratic behavior of the money supply.

**Post-Keynesian Extensions** To Keynes's list of anti-quantity theory arguments, neo-Keynesian economists writing in the inflationary environment of the post-World War II period added others. One was the view that inflation is predominantly a cost-push phenomenon associated with union bargaining strength, monopoly power, administered or mark-up pricing policies, and other non-monetary institutional forces that contribute to autonomous increases in labor and other factor costs. Another was the view, espoused by "cheap-money" advocates, that expansionary monetary policy could be used to peg interest rates at low levels, thus minimizing the interest burden of the public debt. An alternative version of the same argument, associated with the Phillips Curve approach to policy questions, held that monetary policy could help peg the unemployment rate at permanently low levels. These latter two arguments conflict with the neutrality proposition that holds that money can have no permanent influence on real variables.

Perhaps the strongest anti-quantity theory views, however, were those contained in the Radcliffe Com-

mittee's revival and restatement of the Banking School's position on the problem of money and near-moneys. Representing the apogee of post-Keynesian skepticism of the relevancy of the quantity theory, the Radcliffe Report concluded (1) that money is a practically indistinguishable component of a virtually continuous spectrum of financial assets; (2) that the velocity of money is a mere arithmetic computation devoid of volitional meaning or economic content; and (3) that attempts to regulate spending via monetary control are inherently futile in a financial system that can economize on money by producing a limitless array of money substitutes. *Liquidity*, or the total of all assets performing some monetary function, was said to be the key determinant of spending. This variable, it was argued, could expand or contract independently of the narrowly-defined money supply. In the Radcliffe view, attempts to reduce inflation via contraction of the money supply could be frustrated by a compensatory increase in money substitutes, which in the equation of exchange would appear as a rise in the velocity of money. To summarize, the Radcliffe view restated, albeit in modern terms, the old Banking School arguments (1) that it is hard to draw the line between money and near-money, (2) that the volume of credit that can be erected on a given monetary base is virtually unlimited, and (3) that the supply of credit is an endogenous variable responding to the demand for it.

**The Monetarist Counterattack** Quantity theorists responded to the Keynesian attack with counter-arguments based on theoretical developments and empirical research. Chief among the theoretical developments contributing to the resurgence of the quantity theory were (1) the theory of the *real balance* or *wealth effect*, and (2) Milton Friedman's reformulation of the quantity theory as a theory of the demand for money.

The theory of the real balance effect was used to demonstrate that money matters, at least in principle, even in the extreme Keynesian case where the interest rate channel is blocked by a liquidity trap and/or an interest-insensitive investment spending schedule. The key assumptions of the analysis were that real balances are a component of wealth and that wealth is an important determinant of consumption and investment spending. According to the real balance argument, prices would fall in a depression, thereby raising the purchasing power of wealth held in money form. The price-induced rise in the real value of cash balances would then stimulate spending directly until full capacity utilization had been attained. As

the wealth effect operated independently of changes in interest rates, closure of the indirect channel could not prevent the restoration of full employment. It was but a short step from the analysis of the price-induced wealth effect to the argument that a rise in real balances and hence spending could be accomplished just as easily via a monetary expansion, thereby proving the potential potency of monetary policy even in a depression.

In sum, the real balance argument weakened the Keynesian attack in several important respects. At the theoretical level, it offered both an avenue of escape from the Keynesian liquidity trap and a means of thwarting the interest inelasticity of the investment spending schedule, thus contradicting the Keynesian doctrine of underemployment equilibrium. Moreover, it cast doubt on the Keynesian view of money as a specific substitute solely for bonds. It created this doubt by emphasizing the relation between real balances and spending, thus suggesting that money was a general substitute for a wide range of goods and services. Finally, it suggested that the Keynesian view of the monetary transmission mechanism was seriously incomplete.

The second important theoretical development was Milton Friedman's restatement of the quantity theory, a reformulation that emphasized two novel features. First, the quantity theory was reinterpreted as a theory of the demand for money rather than as a theory of the determination of the level of prices and nominal income. Second, the essence of the quantity theory was said to be the existence of a stable functional relationship between the velocity of money (or its counterpart, the quantity of real balances demanded) and a small number of independent variables that determine it.

The reader will notice how Friedman's reformulation was designed to rebut many of the Keynesian criticisms. In denying that the quantity theory was a theory of income determination, Friedman freed it from the Keynesian criticism that it assumed full employment. And in stating the quantity theory as a demand-for-money function capable of being empirically tested, Friedman countered the Keynesian contention that the theory was a mere tautology. Finally, Friedman's treatment of velocity as a stable functional relationship refuted the Keynesian arguments (1) that velocity is a mere arithmetic calculation devoid of economic content; (2) that the quantity theory assumes velocity to be constant; and (3) that velocity is an unstable magnitude subject to erratic, unpredictable shifts. In Friedman's formulation, fluctuations in velocity are perfectly consistent with the idea of a stable functional relation,

since those shifts may be caused by changes in the independent variables of the velocity function.

Quantity theorists also attempted to refute Keynesian criticisms with empirical research. Two types of empirical studies were utilized. The first was a reexamination of American financial history, the main contributions being Friedman and Schwartz's *A Monetary History of the United States, 1867-1960* and Cagan's *Determinants and Effects of Changes in the Stock of Money, 1867-1960*. Both volumes amply demonstrated the significant independent role played by money stock changes in U. S. business cycles. One of the main conclusions of the Friedman and Schwartz study was that a rapid and large reduction in the money supply played the dominant causal role in the Great Contraction of the 1930's. This finding led to the criticism that the Keynesian interpretation, which had attributed the depression to a collapse of investment demand, was a misreading of the facts of experience. The Cagan volume demonstrated, as did the Friedman and Schwartz study, that throughout much of U. S. monetary history the supply of money was independently determined. This evidence seemed to refute the Banking School-Radcliffe doctrine that the stock of money is demand-determined.

The second type of empirical research advanced in defense of the quantity theory took the form of statistical tests, conducted in the early 1960's, comparing the predictive accuracy of Friedman's version of the quantity theory against the rival Keynesian income-expenditure theory. In these tests, the quantity theory consistently out-performed the Keynesian theory. Recent studies, however, have cast doubt on the validity of the basic methodology underlying these tests. Hence, the findings should be regarded as inconclusive. At the time, however, the tests seemed to support the quantity theory.

Associated with the resurgence of the quantity theory has been a rise in the monetarist approach to policy problems. The monetarist view contains several elements. It regards monetary policy as having a powerful long-run impact on nominal income. By contrast, it regards fiscal policy as having a negligible and incomes policy a perverse long-run impact on economic activity. Monetarists, furthermore, argue that the quantity of money, rather than the level and structure of interest rates, is the appropriate variable for the monetary authority to regulate. And finally, monetarists hold that the existence of long and variable lags makes it difficult to predict the short-run impact of monetary changes; therefore, discretionary stabilization policy should be abandoned in favor of a rigid rule whereby the

money supply grows at a fixed percentage rate corresponding to the long-term growth rate of real output. These monetarist policy prescriptions have gained increasing recognition in recent years.

### **SURVIVAL OF THE CLASSICAL QUANTITY THEORY IN THE MODERN MONETARIST APPROACH**

This article has sketched the evolution of the quantity theory of money from its fragmentary pre-classical beginnings, through its crystallization and consolidation in classical monetary analysis, and finally to its culmination in the recent rise of monetarism. Among the milestones in this long process of historical development were: (1) Bodin's hypothesis concerning the cause of the 16th century price revolution; (2) the Cantillon-Hume two-fold distinction between (i) equilibrium statics and disequilibrium dynamics, and (ii) the long-run neutrality and short-run non-neutrality of money, both distinctions necessary to an understanding of the causal role of money; (3) the classical economists' application of the theory to policy questions concerning the regulation of the money supply; (4) the mathematical restatement of the theory by Irving Fisher and other neo-classical economists; and, finally, (5) Milton Friedman's recent reformulation of the quantity theory as a theory of the demand for money.

To some extent, however, this evolution has been illusory. Despite the apparent growing sophistication and complexity of the theory, there has been no radical change in its basic tenets since the early 19th century. After having gained firm roots in classical monetary tradition, the fundamental postulates of the theory experienced little subsequent alteration. Since the classical period, most of the improvements in the theory have consisted of its periodic and increasingly rigorous reformulation in order to conform with the latest innovations in economic theorizing or to meet the increasingly severe standards of empirical testing. Examples include (1) Fisher's reformulation of the theory in terms of the equation of exchange, which corresponded to the emerging use of mathematics in neo-classical economic analysis, and (2) Friedman's restatement, which utilized the latest developments in capital theory and incorporated the asset or portfolio approach to the demand for money, and which facilitated statistical estimation and testing.<sup>2</sup> These refinements changed the outward appearance of the theory without altering its underlying postulates.

<sup>2</sup> It should be noted that the balance-sheet or asset-portfolio approach is not of monetarist origin. This approach was first developed by Keynes and J. R. Hicks in the mid-1930's and was elaborated subsequently by James Tobin and others.

The survival of these 19th century monetary postulates serves to link the older with more modern explanations of the quantity theory. It is, therefore, fitting to close the article with a brief comparison of the chief conclusions of the classical and the monetarist versions of the theory. The classical explanation, it will be remembered, stressed: (1) the neutrality of money in long-run equilibrium, (2) the temporary non-neutrality of money in the transition period, (3) the causal role of money in the transmission mechanism, (4) the monetary theory of price movements, (5) long-run proportionality between money and prices, and (6) exogeneity of the money supply. Moreover, the classical policy analysis yielded the additional conclusions that the money supply can be effectively regulated through the control of its note component alone and that time lags render discretionary monetary stabilization efforts destabilizing, thus necessitating the imposition of a fixed rule.

In line with the classical notion of the long-run neutrality of money, monetarists still argue that the long-term expansion path of output is determined by real factors, e.g., resource endowments, technology, and the productivity of labor and capital. It is argued that changes in the money stock can have no long-run impact on these real determinants of output. Consequently, in long-run equilibrium money is merely a veil. Monetarists, furthermore, adhere to the classical doctrine that the real rate of interest is determined by the non-monetary forces of productivity and thrift. They reject the neo-Keynesian view that the monetary authorities can permanently alter the real rate of interest (and thus the pace of capital formation and the growth rate of output) via changes in the money supply. Likewise they reject the neo-Keynesian notion that an expansionary monetary policy can permanently peg the rate of unemployment at low levels. There can be no long-run relation, say the monetarists, between a monetary variable and real variables, such as the interest rate and the rate of unemployment.

Monetarists adhere to the classical doctrine of the temporary short-run non-neutrality of money. They stress that any sudden change in the money supply or its rate of growth will have a significant frictional impact on output, employment, and perhaps the product-mix. As the chief reason for these non-neutral transition effects, they cite the distortion of relative prices owing to the failure of some prices to adjust as fast as others to the monetary change. They point out that prices do not adjust fully and instantaneously to an unanticipated monetary change because it takes time for people to perceive the change



and adapt to it. Eventually, however, actual money and price level changes are fully recognized and future changes in these variables are correctly anticipated. Consequently, actual price changes in all sectors of the economy adjust completely to price experience and price expectations, thus eliminating the temporary distortions to real variables. The same points, of course, have been accepted by quantity theorists since Cantillon and Hume.

As for the active role of money in the transmission mechanism, monetarists stress that money stock changes precede and cause changes in nominal national income. Similar to their Bullionist and Currency School forebears, monetarists view money as the chief source of economic disturbance and as the predominant cause of price level changes. They attribute both the Great Depression of the 1930's and the post-1965 inflation to the erratic behavior of the money supply. Inflation, they state, is always and everywhere a monetary phenomenon. Since money is the main disrupter of economic equilibrium, it follows that proper control of the money supply is the key to reducing inflation and depression.

The active role of money is also stressed in monetarists' theoretical analysis of the monetary adjustment process. Here the motivating force is always seen to be discrepancy between actual and desired real cash balances, i.e., an excess supply of money. If people have more money than they desire, they will spend the excess for assets, including securities, investment goods, and consumption goods. The increased spending eventually leads to higher prices, thereby bringing actual real balances back to their desired level and thus eliminating the initial excess supply of money. The emphasis in the monetarists analysis is clearly on the causal role played by money in the adjustment to new equilibrium.

On the related issue of the transmission mechanism, the monetarist analysis tends to emphasize interest rate effects more than did the classical analysis, which tended to highlight the direct impact of monetary changes on commodity expenditure. Despite a great deal of lip-service paid to the notion of the direct effect, monetarists now acknowledge that the transmission mechanism operates primarily through a complicated portfolio or balance sheet adjustment process involving numerous interest rate channels and affecting a wide range of assets and expenditures. Specifically, the monetarist views monetary changes as generating shifts in the composition of asset portfolios or balance sheets, thereby inducing changes in the prices and yields of existing financial and non-financial assets (including producer and consumer durable goods) relative to the prices of current serv-

ices and new assets. These asset price and yield changes, in turn, generate changes in the demands for service flows and new asset stocks and hence in the prices and/or outputs of the latter items.

Although the monetarist analysis differs from the classical in the role it assigns to the portfolio-adjustment process, it nevertheless agrees with the classical view of the *strength* of the transmission mechanism. Modern monetarists maintain that the linkages connecting money to spending are numerous, thereby permitting the full impact of a monetary change to be transmitted to prices and nominal income. Monetarists argue that the portfolio-adjustment or asset-substitution effects of a monetary change have a powerful influence on spending because they operate over such a wide range of assets and interest rates. It should be noted in passing that the question of the appropriate range of assets and interest rates to be considered in the analysis of the transmission mechanism is a key point in the monetarist-Keynesian controversy over the spending impact of monetary changes. Unlike the monetarist model, Keynesian models tend to concentrate on a narrow range of assets and interest rates. Consequently, the transmission process is forced to go through an extremely narrow channel, thereby choking off some of the spending impact of a monetary change. No such limitation exists in the monetarist model, which concentrates on a wide range of assets and interest rates. In the monetarist analysis, individuals are seen as disposing of their excess money balances over a broad spectrum of existing assets, including bonds, equities, durable producer goods, durable and semi-durable consumer goods.

Modern monetarists also agree with classical quantity theorists on the question of the exogeneity of the money supply. Like the Bullionists and members of the Currency School, monetarists contend that the central bank can exercise effective control over the nominal money stock by controlling a narrowly defined base of high-powered money (currency plus bank reserves). More specifically, monetarists maintain that the supply of money is determined by three distinct variables, including (1) the monetary base (controlled by the monetary authority), (2) the reserve/deposit ratio (determined by the decisions of commercial bankers subject to legal reserve requirements), and (3) the currency/deposit ratio desired by non-bank individuals. The latter two determinants form the sole components of the so-called money multiplier which, when multiplied by the monetary base, yields the money supply. Monetarists contend that the money multiplier forms a fairly stable link between the base and the money

stock, thus permitting the central bank to exercise effective control over the money supply. It is true that the money multiplier itself is not under the direct control of the central bank. Commercial bankers and the public, via their decisions regarding the desired reserve/deposit and currency/deposit ratios, determine the size of the multiplier. But monetarists argue that the money multiplier and its component ratios are sufficiently stable and predictable to permit firm control of the money stock via control of the monetary base.

On other policy issues relating to money control, monetarists echo the views of their 19th century predecessors. Today's monetarists are no less critical of the central bank than were Ricardo and other Bullionist writers of the 19th century Bank of England. Similar to Ricardo, who pointed out that by pegging the interest rate, the central bank may lose control of the money supply, modern monetarists insist that interest rates are a misleading guide to monetary policy. Like the Currency School, which argued that convertibility was an insufficient safeguard to overissue because of time-lags that might bring the central bank's gold reserve near to exhaustion, modern monetarists argue that the existence of long and variable lags in the relation between money, income, and prices, as well as the lack of understanding of those lags, militates against the use of discretionary monetary policy. The effect of these lags, monetarists hold, is to make the short-run response of income to monetary changes erratic and hard to predict. It follows, therefore, that short-run stabilization policy is at best difficult and at worst likely to be perverse and hence should be avoided. Following the example of the Currency School, monetarists advocate that the central bank's discretionary management be replaced by a rule—in this case a rule fixing the annual percentage growth rate of the money stock at a steady figure roughly corresponding to the long-term growth rate of output.

Monetarist doctrine departs from the strict classical quantity theory on at least one major point, i.e., the proportionality postulate.<sup>3</sup> Monetarists do not insist

<sup>3</sup> There are other differences of course. For example, modern monetarists employ a more comprehensive measure of the money supply—defined to include demand deposits as well as notes and coin—than did their classical predecessors. Moreover, modern monetarists also favor flexible foreign exchange rates whereas the Currency School advocated fixed rates.

on a rigidly proportional relationship between monetary changes and price level changes. As previously mentioned, the proportionality postulate follows from the classical assumption of constancy in the quantity of real cash balances demanded by moneyholders. If real cash balances are to remain unchanged following a change in the nominal money stock, then a rise in  $M$  must necessarily be accompanied by an equiproportional rise in  $P$  to keep real balances,  $M/P$ , the same. Unlike classical quantity theorists, however, monetarists do not interpret the quantity of real balances demanded as a numerical constant. Instead, they view it as a stable function of several variables, including wealth, real income, expected real rates of interest (the opportunity cost of money holdings), and the anticipated rate of inflation (the depreciation cost of cash balances). Depending on movements in these variables, the quantity of real balances demanded may alter from time to time. For example, if a monetary injection leads to a rise in the anticipated rate of inflation, the quantity of real balances demanded will fall. People will want to hold a smaller quantity of real balances than before because of the rise in the depreciation cost of money holdings. In consequence, prices will rise in *greater* proportion than the change in the money stock. The greater-than-proportionate rise in  $P$  is necessary to achieve the desired reduction in real balances,  $M/P$ .

Nevertheless, the difference between monetarists and classical quantity theorists on the proportionality question is not very great. Monetarists insist that, under normal conditions, the quantity of real balances demanded is a definite and stable magnitude. Real balances demanded may fall slightly when the money supply increases, or rise somewhat when the money supply falls, or perhaps even undergo some alteration in the absence of monetary change. But these changes are not expected to be very large. In short, while desired real balances are no longer viewed as a constant, they are seen normally as being subject to only very moderate changes. Under such conditions any monetary change will be accompanied by near-proportional change in prices. Practically speaking, therefore, monetarists would probably accept the proposition of near-proportionality between money and prices in the long run.

Thomas M. Humphrey