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, Allan 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.¹

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

¹ 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 confuse 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 1732 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 statics 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 convertible 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 convertible 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 enacting 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-
The criticisms 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 resources are unemployed 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 neutral. 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, money 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 spillover 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 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 overissue, 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 economists argued that the theory of the quantity of money was untenable. They maintained that the amount of money in circulation was determined by the demand for it, and that changes in the money supply were the result of changes in the demand for money. They also argued that the quantity theory was irrelevant to the study of the business cycle, and that the cause of business cycles was to be found in the operation of the real economy, not in the operation of the money market.
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 Antibilionist predecessors, the Banking School contended that currency overissue was impossible as long as banks restricted their loans to self-liquidating commercial or agricultural paper. But the Banking School went further than the Antibilionists, 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 overissue. 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-
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