MONETARY THEORY: NEW AND OLD LOOKS

MONEY, CAPITAL, AND OTHER STORES OF VALUE

By JAMES TOBIN
Yale University

I. Monetary Economics and Rational Behavior

The intellectual gulf between economists' theory of the values of goods and services and their theories of the value of money is well known and periodically deplored. Twenty-five years after Hicks's eloquent call for a marginal revolution in monetary theory [4] our students still detect that their mastery of the presumed fundamental, theoretical apparatus of economics is put to very little test in their studies of monetary economics and aggregative models. As Hicks complained, anything seems to go in a subject where propositions do not have to be grounded in someone's optimizing behavior and where shrewd but casual empiricisms and analogies to mechanics or thermodynamics take the place of inferences from utility and profit maximization.

From the other side of the chasm, the student of monetary phenomena can complain that pure economic theory has never delivered the tools to build a structure of Hicks's brilliant design. The utility maximizing individual and the profit maximizing firm know everything relevant about the present and future and about the consequences of their decisions. They buy and sell, borrow and lend, save and consume, work and play, live and let live, in a frictionless world; information, transactions, and decisions are costless. Money holdings have no place in that world, unless possession of green pieces of paper and yellow pieces of metal satisfies some ultimate miserly or numismatic taste. Wealth, of course, has the reflected utility of the future consumption it commands. But this utility cannot be imputed to money unless there are no higher yielding assets available. As Samuelson has pointed out [4, pages 122-24], in a world of omniscient households and firms, dealing in strictly perfect markets, all vehicles of saving in use must bear the same rate of return. If "money" bears that yield, wealth-holders will be indifferent between money and other stores of value—the demand for money will be indeterminate. If money fails to yield the going rate, no one will hold it. Even though money is required as a medium of exchange, transactors will suffer no cost or inconvenience by holding more lucrative assets at all times except the negligible microseconds before and after transactions.
The general sources of the "utility" of money have, of course, long been clear to monetary theorists. Lavington [9] and Pigou [13], for example, imputed to money a rate of return varying inversely with the size of money holdings relative to the transactions needs and total wealth of the holder. This return stands for the convenience and economy of having wealth readily available as means of payment, as well as the safety of money compared with other stores of value. The only alternative asset that these elders of the Cambridge School explicitly envisaged was capital investment. "This proportion \( k \) depends upon the convenience obtained and the risk avoided through the possession of [money], by the loss of real income involved through the diversion to this use of resources that might have been devoted to the production of future commodities. . . . \( k \) will be larger the less attractive is the production use and the more attractive is the rival money use of resources. The chief factor upon which the attractiveness of the production use depends is the expected fruitfulness of industrial activity" [13, pages 166, 168]. In short, an individual adjusts his money holding so that its marginal imputed return is equal to the rate available to him in capital investment. Paradoxically the Cambridge tradition did not build on these ideas of liquidity preference. Instead of being systematically related to the profitability of investment and to other variables affecting the rational calculations of wealth owners, the demand for money became a constant proportion of income. Marshall [11, page 47] had explicitly mentioned wealth as well as income, but somehow wealth was dropped from the tradition. \((k\) is not the only instance in English economics where a variable coefficient left unprotected by functional notation has quickly evolved into a constant in everyday use.) Hicks's prescription for monetary theory in 1935 was in much the same spirit as the approach of Lavington and Pigou. His strictures were nonetheless timely; the spirit of the original Cambridge theory had become obscured by the mechanical constant-velocity tradition.

Recent developments in economic theory have greatly improved the prospects of carrying out Hicks's "simplifying" suggestions and deriving rigorously the imputed return or marginal utility of money holdings in relation to their size. In the past decade theory has begun a systematic penetration of the murky jungle of frictions, market imperfections, and uncertainties. The theory of optimal inventory holdings, for example, shows how transactions and delivery costs must be balanced against interest and carrying costs. Applied to inventories of cash, the theory gives precision to the relation of cash holdings to the volume of nonfinancial transactions, the costs of asset exchanges, and the yields available on alternative assets [1] and [17]. A parallel de-
development has been the theory of choices involving risk. Applied to the
general strategy of portfolio selection, the theory of risk aversion ex-
plains how money may find a place in a rationally diversified portfolio
[10] and [18].

The new tools are constructing a bridge between general economic
theory and monetary economics. More than that, they give promise at
last of a general equilibrium theory of the capital account. Such a the-
ory would explain both the balance-sheet choices of economic units as
constrained by their net worths and the determination of yields in
markets where asset supplies and demands are balanced. What charac-
teristics of assets and of investors determine the substitutabilities or
complementarities among a set of assets? Among the relevant proper-
ties with which the theory must deal are: costs of asset exchanges;
predictability of real and money asset values at various future dates;
correlations—positive, negative, or zero—among asset prospects; li-
quidity—the time it takes to realize full value of an asset; reversibility
—possibility and cost of simultaneously buying and selling an asset;
the timing and predictability of investors' expected needs for wealth.

In a world of financial assets and well-developed capital markets,
Keynes [7, pages 166 and 168, pages 140-41] was right in perceiving the
tactical advantage to the theorist of treating separately decisions de-
termining total wealth and its rate of growth and decisions regarding
the composition of wealth. A theory of the income account concerns
what goods and services are produced and consumed, and how fast
nonhuman wealth is accumulated. The decision variables are flows. A
theory of the capital account concerns the proportions in which various
assets and debts appear in portfolios and balance sheets. The decision
variables are stocks. Income and capital accounts are linked by ac-
counting identities—e.g., increase in net worth equals saving plus capi-
tal appreciation—and by technological and financial stock-flow rela-
tions. Utilities and preference orderings attach to flows of goods and
services; the values of stocks are entirely derivative from their ability
to contribute to these flows. Some stock-flow relationships are so tight
that this distinction is pedantic: the only way an art collector can ob-
tain the flow of satisfactions of owning a particular chef d'oeuvre is to
own it. But there is a vast menu of assets whose yields are generalized
purchasing power, nothing less or more—investors do not have intrinsic
preferences among engravings of security certificates.

II. The Capital Account in Aggregative Models

Strictures on the Need for Explicit Assumptions. Aggregative models
of the income account reduce the dimensions of general equilibrium
theory, purchasing definiteness in results at the risk of errors of aggre-
Commodities, prices, and factors of production are limited to one or two. For similar reasons, it is fruitful to limit the number of assets in aggregative theory of the capital account.

The first requisite of a theory of wealth composition is that decisions about assets and debts must, in the aggregate as for the individual, add up to the net worth of the moment, neither more nor less. Monetary theory needs to specify explicitly what forms the nonmonetary parts of wealth can take. Many confusions and disagreements can be traced to ambiguities and differences in assumptions about the nature of wealth. A theory should state the menu of assets assumed available, specifying which are components of net private wealth (capital stock plus government debt) and which are intermediate assets (private debts). Moreover, the independent interest rates in an aggregative system should be enumerated. An independent rate is one that is not tied to another yield by an invariant relationship determined outside the system; e.g., by a constant risk differential.

The means of payment of a country—at least in part governmental in origin—are generally demand “debts” of the central government. But there are also means of payment of private manufacture; indeed it is possible to imagine a pure credit economy without government debts of any variety, where all means of payment are private debts backed by private debts. Likewise it is possible to imagine a wholly nonmonetary public debt.

Monetary discussions suffer from confounding the effects of changing the supply of means of payment with the effects of changing the net value of private claims on the central government. The second kind of change takes time and requires private saving, absorbed in fiscal deficit, or dissaving equal to fiscal surplus. The first type of change can be accomplished instantaneously by exchanges of assets. When an author proposes to discuss the effects of changing the supply of money, is he imagining aggregate net worth to change simultaneously by the same amount? Effects that are due to increases of private wealth in the form of government debt should not be attributed to money per se. Sometimes we are asked to imagine that everyone wakes up to find his cash stock has doubled overnight and to trace subsequent adjustments. This mental experiment is harmless and instructive, provided its results are not considered indicative of changes in money supply engineered by normal central bank procedures. The overnight miracle increases equally money stocks and net worth; the gremlins who bring the money are not reported to take away bonds or IOU’s. The repercussions are a mixture of effects: partly those of an unanticipated increase in net worth in the form of assets fixed in money value (as if the gremlins had brought bonds instead); partly those of an in-
crease in the supply of means of payment relative to transactions needs and to other assets. The theory of real balance effect [12] is at the same time much more and much less than the theory of money.

Established procedure in aggregative model building is to specify the quantity of money, \( M \), as an exogenous variable determined by the "monetary authorities." The practice is questionable when part of the money supply is manufactured by private enterprise. Banks are not arms of government. The true exogenous variables are the instruments of monetary control: the quantity of demand debt available to serve as primary bank reserves, the supplies of other kinds of government debt, required reserve ratios, the discount rate. Once these instrument variables are set, the interaction of bank and public preferences determines the quantity of money. No doubt a skillful central bank can generally manipulate its controls to keep \( M \) on target, but part of the job of monetary theory is to explain how. A theory which takes as data the instruments of control rather than \( M \), will not break down if and when there are changes in the targets or the marksmanship of the authorities.

**Two Models, One Keynesian and One Not.** The assets of a formal model of Keynes's *General Theory* [7] appear to be four or possibly five in number: (1) government demand debt, serving either as means of payment or as bank reserves, (2) bank deposits, (3) long-term government bonds, (4) physical capital, i.e., stocks of the good produced on the income-account side of the model, and possibly (5) private debts, serving along with bonds (3) and demand debt (1) as assets held by the banking system against its monetary liabilities (2). Net private wealth is the sum of (1), (3), and (4).

Though there are four or five assets in this model, there are only two yields: the rate of return on money, whether demand debt or bank deposits, institutionally set at zero, and the rate of interest, common to the other two or three assets. For the nonmonetary assets of his system, Keynes simply followed the classical theory of portfolio selection in perfect markets mentioned above; that is, he assumed that capital, bonds, and private debts are perfect substitutes in investors' portfolios. The marginal efficiency of capital must equal the rate of interest.

Keynes did not, of course, envisage literal equality of yields on consols, private debts, and equity capital. Indeed, he provides many perceptive observations on the sources and cyclical variations of the expectations and risk premiums that differentiate market yields. But in given circumstances these differentials are constants independent of the relative supplies of the assets and therefore inessential. Once one of the rates is set, the others must differ from it by appropriate allowances for risk and for expectations of price changes.
Thus Keynes had only one yield differential to explain within his theoretical model: the difference between the zero yield of money and the interest rate. This differential he explained in his theory of liquidity preference, which made the premium of bond yields above money depend on the stock of money relative to the volume of transactions and, presumably, aggregate wealth. Keynes departed from the classical model of portfolio choice and asset yields to explain money holdings, applying and developing an innovation borrowed from his own *Treatise* [8, pages 140-44, 248-57], a rate differential that depends systematically on relative asset supplies.

Post-Keynesian aggregative theorists, whether disciples or opponents or just neutral fanciers of models, have stuck pretty close to the Keynesian picture of the capital account. For example, Patinkin [12] explicitly includes all the assets listed above, and no more, in his most comprehensive model. Like Keynes, he has only one interest rate to determine. His difference from Keynes is his real balance effect.

As Hicks [5], Kaldor [6], and others have pointed out, there are apparently no short-term obligations of fixed money value in the Keynesian scheme. Recognition of these near-moneys would add one asset category and a second interest rate to the Keynesian model of the capital account. Transactions costs become the major determinant of the money-short rate differential, and considerations of speculation and risk for investors of different types affect the size and sign of the short-long differential.

An entirely different monetary tradition begins with a two-asset world of money and capital and ignores to begin with all closer money substitutes of whatever maturity. Significantly, the authors of the Cambridge tradition, as mentioned above, regarded direct capital investment as the alternative to money holdings. Why did they fail to carry into their monetary theory the clear inference that the demand for money depends not only on the volume of transactions but also on the yield of capital? Perhaps the best guess is that for these economists the yield of capital was in the short run a constant, explained by productivity and thrift. Money balances were adjusting to a rate already determined, not to a rate their adjustment might help to determine.

On its own logic, therefore, the constant-velocity approximation is of little applicability in models where the rate of return on capital is variable. It is not applicable to cyclical fluctuations, where variations of employment affect the productivity of the given capital stock. It is not applicable to secular growth, if capital deepening or technological change alters the yield of capital.

Neither is the constant-velocity assumption applicable where money substitutes other than capital are available and have endogenously
variable yields, for then the demand for money would depend on those yields. Paradoxically, the model of greatest popularity in everyday analysis of monetary policy really has no room for monetary policy per se. In the two-asset, money-capital economy there are no assets which the central bank and the banking system can buy or sell to change the quantity of money.

What is the mechanism by which a change in the quantity of money brings about the proportional change in money income that constant-velocity theory predicts? Sometimes the mechanism as described seems to assume a direct relationship between money holdings and spending on income account: When people have more money than they need, they spend it. It is as simple as that. Patinkin [12, Chapter 8] rightly objects that spending on income account should be related to excess wealth, not excess money. If the mechanism is a real balance effect, then it works only when new money is also new private wealth, accumulated by the public as a result of government spending financed at the printing press or the mint.

A mechanism more in the spirit of the arguments of Lavington, Pigou, and Hicks is that owners of wealth with excess money holdings seek to restore the balance of their capital accounts. Trying to shift from money to capital, they bid up the prices of the existing capital stock; and since new capital goods and old must bear comparable prices, prices also rise in commodity markets. The process ends when, and only when, money incomes have risen enough to absorb the new money into transactions balances. The real rate of return on the capital stock remains unchanged.

This mechanism can apply to increases in $M$ due to expansion of bank lending—with private debts added to the menu of assets—as well as to increases associated with net saving. One aspect of the mechanism is then the process of which Wicksell [19] gave the classical description. Banks expand the money supply by offering to lend at a rate—the market rate—lower than the yield of capital—the natural rate. Excess demand for capital by new borrowers bids up capital values, with the repercussions already described. Whether this process has an end or not depends on whether the banks’ incentive to expand is extinguished by proportionate increases of money supply, money income, and prices. For a pure credit economy, where all means of payment are based on monetization of private debts, this model produces no equilibrium. The end to the Wicksellian process depends on banks’ needs for reserves, whether enforced by legislation or by their own transactions and precautionary motives.

I have presented a modern version of a two-asset, money-capital economy in [16]. Money and government debt are one and the same,
and there are no private debts. The proportions in which owners of wealth desire to split their holdings between money and capital depend upon the volume of transactions and on the rate of return on capital. The yield of capital is not a constant, as it seems to be in the Cambridge model, but depends on the capital intensity of current production. The differential between the yield of capital and that of money depends on the relative supplies of the two basic assets; the liquidity preference mechanism is applied to a money-capital margin rather than a money-securities margin. The price level adjusts the relative supplies to the portfolios investors desire, given the ruling marginal productivity of capital. This portfolio adjustment is like the mechanism of response to increase in the quantity of money described above for the constant-velocity model; but here it does not necessarily maintain the same velocity or the same yield of capital. A real balance effect on consumption can be added if desired.

A trivial extension of the money-capital model is to include other kinds of government securities, on the assumption that given certain constant rate differentials they are perfect portfolio substitutes for money proper. Then "money" in the model stands for the entire government debt, whether it takes the form of media of exchange or money substitutes. The differential between the return on capital and the yield of any government debt instrument is determined by the relative supplies of total government debt and capital.

By a similar extension private debts could be added to the menu of assets, again with the proviso that they are perfect substitutes for government debt instruments but not for capital equity. This addition does not change the requirement of portfolio balance, that the net private position in assets of fixed money value stands in the appropriate relationship to the value of the capital stock.

Thus extended, the money-capital model winds up with the same asset menu as the Keynes-Patinkin model. Each has only one interest differential to be explained within the model. But there is a vast difference. The Keynes-Patinkin model assumes that all debt instruments are perfect substitutes for capital. The interest rate to be explained is the rate common, with the appropriate constant corrections, to all assets other than money itself. What explains this rate is the supply of money relative to transactions requirements and to total wealth. Monetary policy, altering the demand debt component of government debt, can affect the terms on which the community will hold the capital stock. Expansion of the real value of unmonetized debt cannot do so, although in Patinkin's version it can influence the level of activity via the real balance effect on current consumption. The money-capital model, in contrast, casts debt instruments on the side of money and focuses at-
Attention on the relationship between the total real value of government
debt, monetized or unmonetized, and the rate of return the community
requires of the capital stock. It contains no role for monetary policy;
only the aggregate net position of the public as borrowers and lenders
is relevant, not its composition.

The two models give different answers to important questions. Does
retirement of government long-term debt through taxation have expan-
sionary or deflationary consequences? The question refers not to the
temporary multiplier-like effects of the surplus that reduces the debt—
these are of course deflationary—but to the enduring effects, through
the capital account, of having a smaller debt. The instinctive answer of
economists schooled in the Keynesian tradition is "expansionary." The
supply of bonds is smaller relative to the supply of money; the rate of
interest goes down, and investment is stimulated until the marginal
efficiency comes down correspondingly. The answer of the money-capi-
tal model is, as indicated above, "deflationary." The assumed substi-
tutability of bonds and money will keep the bond rate up. The decline
in the government debt component of net private wealth means that
investors will require a higher rate of return, or marginal efficiency, in
order to hold the existing capital stock.

Granted that both models are oversimplified, which is the better
guide to instinct? Are long-term government debt instruments a better
substitute for capital than they are for short-term debt and money?
Reflection on the characteristic properties of these assets—in particular
how they stand vis-à-vis risks of price-level changes—surely suggests
that if government securities must be assimilated to capital or money,
one or the other, the better bet is money.

Towards a Synthesis. A synthesis of the two approaches must, of
course, avoid the arbitrary choices of both, abandoning the convenience
of assuming that all assets but one are perfect substitutes. The price of
this advance in realism and relevance is the necessity to explain not
just one market-determined rate of return but a whole structure. The
structure of rates may be pictured as strung between two poles, an-
chored at one end by the zero own-rate conventionally borne by cur-
rency (and by the central bank discount rate) and at the other end by
the marginal productivity of the capital stock. Among assets that are
not perfect substitutes, the structure of rates will depend upon relative
supplies. In general, an increase in the supply of an asset—e.g., long-
term government bonds—will cause its rate to rise relative to other
rates, but less in relation to assets for which it is directly or indirectly
a close substitute—in the example, short-term securities and money—
than in relation to other assets—in the example, capital.

In such a synthesis, monetary policy falls in proper perspective. The
quantity of money can affect the terms on which the community will hold capital, but it is not the only asset supply that can do so. The net monetary position of the public is important, but so is its composition.

One lesson of the simple money-capital model should be retained. The strategic variable—the ultimate gauge of expansion or deflation, of monetary tightness or ease—is the rate of return that the community of wealth-owners require in order to absorb the existing capital stock (valued at current prices), no more, no less, into their portfolios and balance sheets. This rate may be termed the supply price of capital. If it is lower than the marginal productivity of capital, there will be excess demand for capital, stimulating increases in prices of capital goods and additions to the stock. If the supply price of capital is higher than its marginal productivity, demand for capital will be insufficient to absorb the existing stock; its valuation will tend to fall, discouraging production of new capital goods. The effects of deviation of supply price of capital from the marginal productivity of the existing stock are similar to those of discrepancies between Wicksell's market and natural rates.

In assessing policy actions and other autonomous changes, there is really no short-cut substitute for the supply price of capital. As the example of long-term debt retirement illustrates, the Keynesian interest rate, the long-term bond rate, can be a misleading indicator. Events that cause it to fall may cause the supply price of capital actually to rise. Another example of error due to concentration on the long-term bond rate is the following Keynesian argument: Expectation of a rise in the interest rate leads to liquidity preference and keeps the current interest rate high; a high interest rate discourages investment. However, what the marginal efficiency of capital must compete with is not the market quotation of the long-term rate, but that quotation less the expected capital losses. If the fact that the rate so corrected is close to zero causes substitution of money for bonds, should it not also cause substitution of capital for bonds?

If the long-term bond rate is an inadequate substitute for the supply price of capital, the same is true of another popular indicator: the quantity of money. The modern quantity-of-money theorist [2] (to be distinguished from the ancient quantity-theorist-of-money, who actually was a believer in the constancy of velocity), holds that virtually everything of strategic importance in the capital account can be studied by focusing on the supply and demand for money. This view, though seemingly endorsed by Shaw [15], has been persuasively opposed by Gurley and Shaw [3]. As they point out, it is not hard to describe events and policies that raise the supply price of capital while leaving
the quantity of money unchanged or even increasing it. Why concentrate on variables other than those of direct central interest?

How far to go in disaggregation is, as always, a matter of taste and purpose; it depends also on the possibilities of empirical application and testing. A minimal program for a theory of the capital account relevant to American institutions would involve: (1) four constituents of net private wealth: government demand debt, government short debt, government long debt, and capital stock; (2) two intermediate assets: bank deposits and private debts; (3) two institutionally or administratively fixed interest rates: zero on bank deposits and demand debt, and the central bank discount rate; (4) four market-determined yields: the short-term interest rate, the long-term interest rate, the rate on private debts, and the supply price of equity capital.

In this model, the quantity of demand debt is divided between currency held outside banks and the net (unborrowed) reserves of banks. Required reserves depend on the volume of deposits. If required reserves exceed net reserves, banks must borrow from the central bank at the discount rate. The disposable funds of banks are their deposits less their required reserves. These are divided among net free reserves (net reserves less required reserves), short governments, long governments, and private debts in proportions that depend on the discount rate, the short rate, the long rate, and the private loan rate. The nonbank public apportions net private wealth among currency, bank deposits, the two kinds of interest-bearing government debt, private debt to banks (a negative item), and capital equity. All the yields except the discount rate are relevant to the public’s portfolio choices. When the wealth constraints are allowed for, there are four independent equations in this system; e.g., a balance equation for each constituent of net private wealth. These equations can be used to find the four endogenous yields. The solution for the yield of capital is its supply price. There is equilibrium of the whole system, which would include also equations for the income account, only if the solution for the supply price of capital coincides with the marginal productivity of the existing stock.

References