THE THEORY
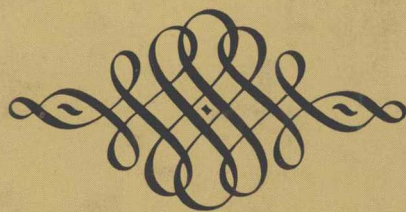# OF MONEY

JÜRG NIEHANS

JÜRG NIEHANS

The Theory of Money

The Johns Hopkins University Press
Baltimore and London

PREFACE

This book is meant to be an introduction into the pure theory of money and monetary policy. Its main topic is the interaction of the money stock with other assets and the flows of goods and services. The book is a "treatise" in the sense of a systematic, comprehensive treatment of the principles of its subject matter. In the monetary field there have appeared no such treatises since Patinkin's *Money, Interest and Prices*, first published in 1956. Since then monetary theory has changed considerably. The present book tries to provide a new synthesis of these developments.

Like other treatises of this sort, this book combines features of both a textbook and a research monograph without claiming to be either. It is similar to a research monograph inasmuch as it includes original contributions that cannot be considered to be "standard analysis." It is similar to a textbook inasmuch as it gives a systematic discussion of the subject, including introductory and expository material and references to the literature.

To achieve a comprehensive treatment in a volume of moderate size, the analysis is limited to fundamental aspects that promise to be of general and lasting significance over a wide range of institutions, countries, and historical periods. In comparison with the two volumes of Keynes' *Treatise on Money*, the present book includes only the "pure theory" and leaves out the "applied theory" of money.

Money is considered in both its microeconomic and macroeconomic aspects. The analysis thus extends from the demand for money by individuals to monetary policy. International aspects, however, are, with a few exceptions, left out, since they would require a separate volume. The theory of monetary growth is also treated only briefly, because after a rapid development in the late 1960s it has progressed little since Stein's *Money and Capacity Growth*. At the microeconomic level the approach is characterized by the decisive role assigned to transaction costs. At the macroeconomic level the main characteristic is the use of a general stock/

flow approach, which includes Keynesian, monetarist, and other positions as special cases.

Chapter 1 reviews the tradition of neoclassical monetary theory by describing how it approached the interrelated issues of the services of money, of its neutrality, and of the integration of monetary and value theory. The subject of chapters 2–4 is the amount of cash balances an individual agent wishes to hold under given market conditions, depending on such factors as real income, commodity prices, interest rates, transaction costs, inflation, and uncertainty. No direct utility of money will be postulated; whatever utility money may have is reflected from the goods money can buy. The key to such an indirect-utility approach is the concept of transaction costs, with factors like speculation and uncertainty appearing as modifying influences.

Chapters 5–7 extend the perspective from individual equilibrium to general equilibrium. It is shown in chapter 5 how market prices and interest rates are determined by the interaction of individual agents, which leads to the problem of the welfare effects of continuing inflation and of the so-called optimal money supply. Chapter 6 raises the fundamental question of the relative advantages of monetary exchange as compared to barter, and of the conditions under which monetary exchange will emerge. While these problems are concerned with money as a medium of exchange, chapter 7 considers the role of money as a medium of account, including the topics of money illusion, deflators, and escalator clauses.

While the first part of the book is mostly about the demand for money, the second part is mostly about the money supply. Commodity money, with the important special case of the gold standard, is analyzed in chapter 8. Chapter 9 investigates the supply of money by commercial banks, developing the argument from the microeconomics of the individual bank to the macroeconomics of the banking system. The following two chapters, covering the traditional subjects of monetary macroeconomics, consider the effect of government-created fiat money on prices, interest rates, employment, and output. While chapter 10 uses a comparative-static approach, describing the adjustments to monetary policy as a sequence of distinct stages, chapter 11 views the adjustment as a dynamic process along a Phillips curve. Chapter 12, finally, discusses the role of the central bank in the economic system and the principles which might guide its monetary policy.

Each chapter opens with a brief nontechnical summary of its content. By perusing these previews, the reader may thus obtain an introduction to the book as a whole.

The main use of the book will probably be at the level of first-year graduate courses, possibly also in advanced undergraduate courses. In fact, the present manuscript has evolved from lecture notes over a period of more than twelve years. The reader is expected to be familiar with the

basics of general equilibrium analysis. While mathematics is extensively used, the required tools do not go beyond the elements of multivariate calculus and nonlinear programming. Graphic devices are used whenever possible. The emphasis is on economic substance and not on mathematical rigor.

Having worked on this book for a long time, I owe debts to many colleagues, friends, and helpers that are not reflected in the references. Foremost among these are my former colleagues at Johns Hopkins in the field of macroeconomics and monetary theory, particularly Carl Christ, Louis Maccini, and William Poole. Dermot McAleese, as a research assistant, helped me relate my own thinking to the tradition of monetary theory. For chapter 11, Paul McNelis provided much of the background analysis and all of the computing work; his contribution amounts to coauthorship. The National Science Foundation granted financial support for a limited part of the research on various aspects of the demand for cash balances and on the role of the central bank. Libby Pratt, Becky Ford, Joyce Goldberger, and Susan Donaldson helped me make my manuscript, through its many stages, ready for the printer. My largest, though most intangible, debt is to The Johns Hopkins University, which for eleven years provided the fertile environment for my work.

CONTENTS

One

THE NEOCLASSICAL TRADITION

Introduction

The most fundamental question of monetary theory concerns the services or "functions" of money. Why do people use money? What do they use as money? How much money do they wish to hold? How do they use it? Economists have always found it more difficult to analyze the services of a medium of exchange than those of producer and consumer goods. This is because a medium of exchange derives its usefulness from some sort of imperfection or "friction" in the market, while the essentials of value and allocation can be understood on the assumption of perfect or "frictionless" markets. For about a century, economic thinking about these problems has been dominated by what may be called the "neoclassical" tradition,[1] characterized by an effort to incorporate money into the general equilibrium framework without making the underlying market imperfections explicit. While the verbal discussions of monetary services are full of references to those imperfections, the latter leave no trace in the formal models. In the final analysis, money is typically treated *as if* it were one of those goods whose services does not depend on market imperfections. We may call this a metaphorical approach to monetary theory.

What will gradually develop over the following chapters constitutes a departure from the neoclassical tradition in the sense that it offers an explicit analysis of frictions. At the same time it may be regarded as an extension of the neoclassical tradition in the sense that money and exchange are incorporated into a general equilibrium framework. As a background to this development, the present chapter offers a review, necessarily brief, of what the neoclassical tradition in monetary theory was.

From early times up to the present day, three propositions, though not

---

1. It is remarkable how little Samuelson's account (1968) of what he was taught in the thirties differed from what he could have learned from Walras at the turn of the century or from what is being taught in better colleges today. Few areas of economics have seen so little progress over this period.

1

universally accepted, have stood in the center of theoretical discussions about money:

1. The use of money increases the efficiency of the economy.

2. Money is neutral in the sense that an exogenous change in its quantity, once all adjustments have run their course, produces a proportional change in all prices, leaving real phenomena unchanged.

3. While money may be different from other goods in important respects, the difference is one in degree rather than in kind; in principle, it should be possible to explain the value of money by the same analytical apparatus that is used to explain the value of other goods.

Since the dawn of economics there has been, in addition, a fourth proposition, accepted by mercantilists and economists, both classical and neoclassical, and reaffirmed by Keynesians, anti-Keynesians, and contemporary theorists. That proposition says that during *transitional* periods a change in the supply of money may have powerful real effects, resulting in fluctuations of output and employment. The discussion of this proposition will be deferred to the second part of this study (see, in particular, chapter 11); the three equilibrium propositions, however, will now be discussed in turn.

1.1. Money and efficiency

Economists (and laymen) have always felt that the use of a medium of exchange increases the efficiency of an economy. The gain was usually considered to be large. It has both qualitative and quantitative aspects. The qualitative aspects appear when monetary exchange is compared with barter. Classical and neoclassical economists were graphic in describing the "double coincidence of wants" of the hungry tailor and the shivering baker which would be necessary for an exchange in a barter economy and the narrow limitations it imposes on the division of labor.[2] The use of money would increase welfare by freeing exchange from the shackles of the double coincidence of wants. It was recognized that the utility of money is due to this increase in welfare, and we can even find the proposition that the supply of money is at the optimum where the marginal cost of an exchange is the same for direct barter and monetary exchange.[3] However, to this day economic theory has never developed an explicit model of multiperson barter.[4] As a consequence, the welfare gain from monetary exchange could

2. A good example is the first chapter of Jevons' *Money and the Mechanism of Exchange* (1875).

3. See Kinley (1904, p. 131).

4. For a recent contribution see Starr (1972). Starr analyzes the (hardly startling) proposition that indirect barter cannot be worse than direct barter and is generally better. The interesting thing is that a century after Jevons such an elementary proposition is still considered to be in need of analytical elucidation—and rightly so.

never be formally analyzed.[5] In particular, it was not clear whether the transition from barter to monetary exchange could be expected to take place spontaneously under the pressure of market forces whenever it promised welfare gains[6] or whether it required an "invention," combined with persuasion, convention, or compulsion.[7] In a vague sense, the inefficiency of barter was often attributed to friction, and money was compared to the lubricant that would reduce that friction. In a frictionless economy, therefore, monetary exchange would have no advantage over barter. An analysis of the role of money in economic equilibrium would thus require a theory of frictions.[8]

Potential elements of such a theory were often mentioned. Thus it was explained that frictions would primarily appear in the form of transaction (or transfer) and storage (or inventory) costs. Writers of neoclassical textbooks seemed to enjoy describing the burdensome inventory and transaction costs in a barter economy. Money, to use Mill's words, thus became a "contrivance for sparing time and labor."[9] The crucial importance of transaction and storage costs is also implied in the recurring discussions of the qualities that make a commodity suitable as a medium of exchange, for if we learn that efficient money should have properties like portability, indestructibility, homogeneity, divisibility, and cognizability,[10] these virtues are easily recognized as some of the numerous determinants of transaction and storage costs. However, up to the present time it was not possible to advance the analysis beyond the collection of plausible suggestions to the construction of contrasting theories of barter and monetary exchange.

The quantitative aspects of the advantages of monetary exchange appear when, in a fully monetized economy, at given commodity prices, an individual is equipped with gradually rising cash balances. The benefit from these balances would have different sources.[11] First, the holder of money

5. What neoclassical economists liked to call a theory of barter was, in fact, something very different, namely a theory of exchange in a world in which it makes no difference who trades with whom and what is exchanged for what. Patinkin's first chapter (1965) illustrates this point. The statement that neoclassical economics was "really" a theory of barter was usually meant to be a severe criticism; it was, in fact, an exaggerated claim. The valid criticism is that neoclassical economics never advanced from a world in which every commodity could serve as a perfect money to the analysis of genuine barter.

6. This was prominently argued by Carl Menger (1871, pp. 253 ff.).

7. For a recent expression of this view see Kurz (1974).

8. See, for example, Hicks (1935).

9. Mill (1857, 2:9).

10. Jevons (1875, chap. 5).

11. For a representative example see Pigou (1917). It is worth noting that the discussion of the familiar transactions, speculative and precautionary motives for holding money did not make much progress from Cantillon to (and including) Keynes.
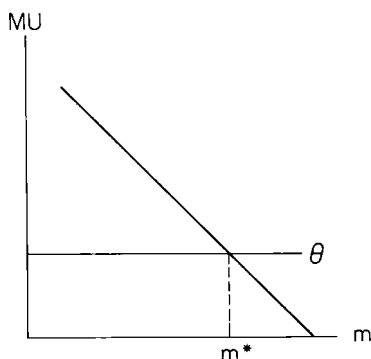
MU

Figure **1.1.1**
The Marginal Utility of Money

would not be constrained to synchronize his purchases and his sales; he could thus schedule his transactions in a more advantageous way. This benefit would, in part, depend on the "payments habits," which were usually regarded as technologically and institutionally determined. Second, cash balances would offer insurance against losses arising either from unexpected needs or from unexpected business opportunities foregone, introducing a precautionary element into the demand for money. Third, the holding of cash might save brokerage and other transaction costs.

The marginal benefit from larger and larger cash balances would presumably decline, ultimately reaching zero where the individual is satiated with money.[12] For given prices, the marginal utility of money can thus be described by a downward-sloping curve (fig. 1.1.1). On the other hand, cash balances might cost something. The cost could be identified as the sacrifice, due to time preference, involved in the holding of wealth instead of consuming it and/or as the income foregone in not holding the wealth in the form of productive capital goods or interest-bearing assets. The individual would try to achieve that level of cash balances which made the marginal utility equal to the marginal cost. If there was indeed a positive marginal cost, the accumulation of cash balances would stop short of satiation.

This argument implies that for given economic conditions each individual has a definite demand for money. In figure 1.1.1 this demand is m^*, which makes the marginal utility of money equal to the marginal cost as expressed (in the absence of capital goods and credit) by the rate of time

12. Pigou quotes Carver: "Some exchanges could scarcely be made at all without the use of money. In these cases the utility of money is very high. . . . Some exchanges could only be made with great difficulty without money, in which cases the utility of money would be considerable. Some exchanges could be made with comparatively little difficulty, in which cases the utility of money would be inconsiderable. And some exchanges could be made as easily without money as with it, in which cases the utility of money will be *nil*" (1917, p. 167).

preference, θ. For given tastes, payments habits, risk, and time preference (but in the absence of credit and capital goods), this demand could be written

$$m = m\,(\bar{x}_1,\cdots\bar{x}_Q;\,p_1,\cdots p_Q;\,\theta),\qquad\qquad(1.1.1)$$

where $\bar{x}_1,\cdots\bar{x}_Q$ represent the given endowment consisting of goods $1,\cdots Q$, $p_1,\cdots p_Q$ are the market prices of these goods and θ expresses the opportunity cost of cash balances. Demand functions of this type could be added over individuals to give an aggregate demand function of the same form.

The demand function for money was written in many special forms. One of these is

$$m = k\sum_{q=1}^{Q} p_q\bar{x}_q = kp\sum_{q=1}^{Q}\frac{p_q\bar{x}_q}{p} = k\,p\,\bar{x},\qquad\qquad(1.1.2)$$

where p is a suitably chosen price index and $\bar{x} = \sum_{q=1}^{Q} p_q\,\bar{x}_q/p$ is a measure of aggregate resources.[13] In the "cash balance approach," which this equation summarizes, the so-called Cambridge k indicates what fraction of the money value of resources, on the average, economic agents wish to hold in the form of cash balances.[14] Alternatively, the equation can be written

$$mv = p\bar{x},\qquad\qquad(1.1.3)$$

where $v = 1/k$. In this variant v can be identified as the (income) velocity of money.[15] It is clear that the cash balance approach and the velocity approach are economically equivalent.[16] While such special forms of the

13. Cf. Pigou (1917, p. 165).

14. A comment on the interpretation of (1.1.2) may help to forestall misunderstandings at this point. If k is interpreted not as the desired, but as the actual, cash balance per dollar worth of resources, the equation is transformed from a demand function into an accounting identity, called the equation of exchange. If we add to the equation of exchange the threefold macroeconomic proposition that (1) m can be changed exogenously, (2) an exogenous change in m has no lasting effect on k, and (3) an exogenous change in m has no lasting effect on $\bar{x}$, we obtain the quantity theory of money (or, more accurately, of prices). It states that an exogenous change in m produces an equiproportionate change in p. The same collection of symbols can thus be interpreted as an accounting identity, as a demand function for money, or as a theory of the price level. The recent tendency to identify the quantity theory of money with a rather general form of demand function for money seems to have originated with Friedman (1956); it is not part of the neoclassical tradition.

15. This form, though known since the early nineteenth century, has found its classical elaboration in the work of Irving Fisher (1911). On the history of the quantity equation and the quantity theory see Marget's massive study (1938), a monument to polemic scholarship. Note that k has the dimension time, while v has the dimension $1/$time or velocity.

16. In an economy with intermediate transactions the volume of transactions is, of course, larger than final output or income. In this case a choice has to be made between an "income approach" and a "transactions approach." This choice relates to a question of substance, while the choice between a cash-balance approach and a velocity approach is a question of exposition.

demand function for money may be useful expository devices, the substance of the neoclassical approach to monetary theory does not depend on them. Equation (1.1.1) is all that is needed.

The demand function for cash balances can easily be incorporated into a general equilibrium system. There is a demand function for each of the Q commodities,

$$x_q = x_q(\bar{x}_1, \cdots \bar{x}_Q; p_1, \cdots p_Q; \theta).$$ (1.1.4)

But the individual can spend no more than he earns, and nothing is thrown away. The aggregate value of commodities demanded thus equals the value of the endowment:

$$\sum_{q=1}^{Q} p_q(x_q - \bar{x}_q) = 0.$$ (1.1.5)

This budget constraint was often called "Say's Law" because it formalizes Say's phrase that "commodities are bought with commodities." The constraint implies that there are only $Q - 1$ independent demand functions. If an individual is subject to just this one budget constraint, this means essentially that commodities can be traded costlessly without limit, provided total purchases equal total sales. If this is interpreted literally, there would be no reason for ever using money or otherwise preferring one medium of exchange to another. Of course, neoclassical monetary theorists would have refused to interpret the constraint so literally, but they were not able to formalize what they thought went on "behind the scenes."

In equilibrium, all markets are cleared:

$$x_q = \bar{x}_q. \qquad (q = 1 \cdots Q - 1)$$ (1.1.6)

If the market is cleared for all but one of the commodities, it is implicitly cleared for the last commodity, because Say's Law has already made aggregate demand over all commodities equal to aggregate supply. If we assume that the economy is equipped with a certain nominal quantity of money, $\bar{m}$, equilibrium also requires

$$m = \bar{m}.$$ (1.1.7)

Disregarding the possibility of multiple solutions and negative prices, these equations determine prices and consumption for all Q commodities and also real cash balances. The quantity of real cash balances in an economy is thus explained jointly with the other variables of the system.

This raises the question whether the quantity of real balances resulting from the interaction of market forces is optimal from the point of view of social welfare. Protagonists of neoclassical theory like Friedman and Samuelson have argued that there is a difference between the private cost of real cash balances and their social cost (for references see chapter 5). The private cost, as perceived by individuals, is equal to the opportunity cost of

real balances as represented in the above model by θ. The social cost, however, is zero: Just by maintaining a lower price level, society can raise real cash balances to any desired level, and a lower price level is costless. The implication is that the market economy tends to be undersupplied with real cash balances. This is what Samuelson called the "nonoptimality of money under laissez faire."

The fundamental problem of the neoclassical general equilibrium system represented by (1.1.1) and (1.1.4) — (1.1.7) is the absence of an explicit account of the factors that make money valuable. Payments habits, risk, and transaction costs do not appear in the system. We do not even know exactly how the marginal utility of cash balances and the demand functions for money are supposed to be derived. While all those picturesque illustrations about the services of money, so dear to the heart of neoclassical writers, are certainly suggestive, they do not amount to analysis.

1.2. The neutrality of money

It is probably the oldest and most familiar proposition of monetary theory that money is demanded not for its own sake but for what it can buy.[17] A neoclassical economist might formalize this statement by saying that the demand function for money (1.1.1) is linear-homogeneous in all prices and can thus be written

$$\frac{m}{p_1} = m\left(\bar{x}_1, \cdots \bar{x}_Q; 1, \cdots \frac{p_Q}{p_1}, \theta\right) \tag{1.2.1}$$

$$= m(\bar{x}_1, \cdots \bar{x}_Q; \pi_1, \cdots \pi_Q, \theta),$$

where p_1 is arbitrarily chosen as numéraire and $\pi_q = p_q/p_1$ are relative prices (with $\pi_1 \equiv 1$). If, for given relative prices, $1/p_1$ is interpreted as the "price of money," the homogeneity postulate can be expressed by saying that the demand curve for nominal cash balances as a function of their "price" is a rectangular hyperbola: The product of m and $1/p_1$ remains constant as p_1 changes.

A neoclassical economist would also argue that the demand for commodities depends only on relative prices and not on the absolute price level. The demand functions (1.1.4) would thus be homogeneous of degree zero,

$$x_q = x_q(\bar{x}_1, \cdots \bar{x}_Q; \pi_1, \cdots \pi_Q; \theta), \tag{1.2.2}$$

while (1.1.5) can be written

$$\sum_{q=1}^{Q} \pi_q(x_q - \bar{x}_q) = 0. \tag{1.2.3}$$

17. Thus Adam Smith wrote: ". . . money can serve no other purpose besides purchasing goods. . . . It is not for its own sake that men desire money, but for the sake of what they can purchase with it" (1776/1904, 1:405).

As a consequence, money can be shown to be neutral in the sense that a change in its nominal quantity changes only the price level, here measured by p_1, while all quantities and relative prices remain unchanged. The nominal quantity of money thus appears to be of no real significance in economic equilibrium.

The typical neoclassical economist would, in fact, feel that the general equilibrium system can be split into a real part—consisting of (1.2.2), (1.2.3), and (1.1.6)—that determines all quantities and relative prices, and a monetary part—consisting of (1.2.1) and (1.1.7)—that determines the price level and real cash balances. It was thus possible, and perhaps convenient, to discuss value and allocation in the first volume of a treatise, while money and the price level would follow in the second volume.[18] The neutrality proposition implied the quantity theory of prices: For given resources and tastes, equilibrium prices would always change in proportion to the exogenous quantity of money. It should be noted that this proposition does not depend on the use of a special demand function for money like one derived from the equation of exchange.

The precise meaning of the neoclassical neutrality proposition was not always clear. First, neoclassical writers sometimes came close to suggesting that *real* cash balances are neutral. Thus John Stuart Mill stated that "the relations of commodities to one another remain unaltered by money," and "things which by barter would exchange for one another will, if sold for money, sell for an equal amount of it" (1857, 2:9 ff.). Taken literally, this says that the economy is indeed neutral with respect to real cash balances; relative prices are the same in a monetary economy as they would be in the absence of money.

This interpretation poses a dilemma. If money does not affect any real variables it cannot increase efficiency, and if it really increases efficiency it cannot leave relative prices unaffected. Typically, classical and neoclassical economists relied on their metaphors to dispose of this point. Money as a "veil" would not be quite satisfactory, since it does not suggest any real significance, but only obfuscation of the underlying realities. Money as a "lubricant" would be just about right, since oil permits the machine to run without being a part of it.[19] Of course, the strict interpretation of Mill's statement is patently invalid (and Mill perhaps did not mean what he said). Neoclassical economists never seriously maintained that money is a "veil" in the sense that it does not affect the shape of what it is hiding.

18. The order could not be reversed; while relative prices could be discussed independently of the absolute price level, the price level could only be determined once relative prices were known.

19. See Jevons (1875, p. 15), Marshall (1923, p. 38), and Wicksell (1934/35, 2:5). Samuelson updates the metaphor to that of money as a catalyst (1968, p. 3).