ROUTLEDGE

money

What it is, how it's created, who gets it, and why it matters

SERGIO FOCARDI

ECONOMICS IN THE REAL WORLD

Money

What It Is, How It's Created, Who Gets It, and Why It Matters

Sergio M. Focardi

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

2018

Money

By enabling the storage and transfer of purchasing power, money facilitates economic transactions and coordinates economic activity. But what is money? How is it generated? Distributed? How does money acquire value and that value change? How does money impact the economy, society?

This book explores money as a system of "tokens" that represent the purchasing power of individual agents. It looks at how money developed from debt/credit relationships, barter and coins into a system of gold-backed currencies and bank credit and on to the present system of fiat money, bank credit, near-money and, more recently, digital currencies. The author successively examines how the money circuit has changed over the last 50 years, a period of stagnant wages, increased household borrowing and growing economic complexity, and argues for a new theory of economies as complex systems, coordinated by a banking and financial system.

Money: What It Is, How It's Created, Who Gets It and Why It Matters will be of interest to students of economics and finance theory and anyone wanting a more complete understanding of monetary theory, economics, money and banking.

Sergio M. Focardi is Professor of Finance and researcher at the Léonard de Vinci Pôle Universitaire, Paris-La Defense, France. He is a member of the editorial board of the *Journal of Portfolio Management* and has authored numerous articles and books on mathematical finance, including a series of monographs for the CFA Institute Research Foundation, including the recent *Equity Valuation: Science, Art, or Craft?*

Economics in the Real World

For a full list of titles in this series, please visit https://www.routledge.com/Economics-in-the-Real-World/book-series/ERW

1. Education Is Not an App
The Future of University Teaching in the Internet Age
Jonathan A. Poritz and Jonathan Rees

2. The Privileges of Wealth
Rising Inequality and the Growing Racial Divide
Robert B. Williams

3. Neuroliberalism
Behavioural Government in the Twenty-First Century
Mark Whitehead, Rhys Jones, Rachel Lilley, Jessica Pykett, and Rachel Howell

4. The Rise of Big Government
How Egalitarianism Conquered America
Sven R. Larson

5. Money
What It Is, How It's Created, Who Gets It, and Why It Matters
Sergio M. Focardi

Acknowledgments

The author thanks Caroline L. Jonas who has read and commented on this book from its conception to its completion and has made a significant contribution to the historical notes.

Contents

Exhibits

Boxes

Introductory Remarks

This book is intended for students of economics and finance theory as well as for people with knowledge of and interest in economics and, in particular, in the role of money in economies and societies. The question of the creation and allocation of money is important for understanding the role of money in advanced economies and, in particular, how it might originate financial or economic instability and inequality – a source of social instability. The question of the creation and allocation of money is part of a more general theory of money and, in particular, of how money interacts with the real economy.

The "theory of money" is not a natural law. Money, its nature and functioning, depends on the institutions and rules of an economy and society at large and evolves together with them. Nonetheless, we will discuss the theory of money within the framework of scientific methods.

What do scientific methods bring to our understanding of money? In a nutshell, the scientific method brings a clear underpinning of concepts and variables, defining them *operationally*, in terms of observations. Economic concepts are often rooted in intuition. Without scientific methods, economic discussions can become obscure attempts to refine and quantify intuition. But science is neither a refinement nor a quantification of intuition. Modern science is based on theories that explain observations. Indeed all concepts in science are ultimately defined in terms of observations.

Historically, science did indeed have its origins in intuition. The scientific endeavor started as an attempt to explain and quantify concepts that were well-known intuitively. But the development of science required concepts firmly rooted in observations, without the need to postulate the intuition of concepts. Consider the notion of force. Everyone has an intuition of "force" that derives from the physical efforts we experience when we move a heavy object. But the physical concept introduced in physics by Isaac Newton in his *Principia* (1687) is an abstraction that cannot be considered an explanation of the intuitive notion of force. Note that the abstract concept of force introduced by Newton makes physical sense only because we can compute forces such as gravitational force.

In modern science there is no place for a priori concepts. Modern science is based on formulating theories that are empirically validated. All concepts of modern science are ultimately defined in terms of observations. Defining concepts in terms of observations proved to be a critical step in modern physics. Consider, for example, the formulation of the Theory of Relativity. Based on the constancy of the speed of

light, the Theory of Special Relativity proved that simultaneity depends on the observer.

However, economics has many theories that are based on a priori concepts not linked to observations. Examples include utility, rational expectations, and price levels. Utility is a concept that is mathematically well-defined but in practice impossible to observe – either directly or indirectly through economic theory. The concept of rational expectations is another mathematical concept with no correspondence to any empirical observation of real-world agents. As for price level, this again is a concept that is not uniquely defined. Each price index, such as the Laspeyres index or the Paasche index (discussed in Section 6.3), defines a different concept of price level. The notion that these indices are approximations of a true index is simply false. Any theory that uses the concept of inflation is therefore a theory relative to the inflation measurement process. This affects both the theory of growth and monetarism.

Science does not explain everything: so what?

Scientific theories respond *in toto* to the empirical test; partial statements cannot be tested. This point was made by, among others, the influential American philosopher and logician Willard van Orman Quine (1953) who wrote:

> No particular experiences are linked with any particular statements in the interior of the field, except indirectly through considerations of equilibrium affecting the field as a whole. If this view is right, it is misleading to speak of the empirical content of an individual statement – especially if it is a statement at all remote from the experiential periphery of the field.
>
> *(p. 43)*

Consider, for example, the well-known law $F = ma$ linking force, mass, and acceleration in classical dynamics. This law is empirically meaningless unless we know how to write laws for force and mass. In Newtonian dynamics this link is given by laws such as the law of gravitation. In practice, in modern science, observations are made through complex instruments whose functioning depends on the entire theory.

Empirical validation – necessarily based on a finite non-exhaustive number of observations – is *always* hypothetical. Scientific theories are only hypotheses that explain all known observations within their domains and are not in contradiction with any known fact. However, there is no guarantee that new observations will not contradict existing theories. In the history of science, many observations contradicted existing theory: these observations led to new, more powerful theories.[1]

Even if scientific theories explain all known observations, it is clear that science does not explain everything. There are reasons for the limitations of science. Science deals with structures. Physical properties are explained in terms of structures such as the configuration of atoms, molecules, or crystals. But some structures are so complex that they cannot be computed. In addition, in some domains, structure alone is not sufficient and we do not even have the concepts to formulate the correct qualitative questions. For example, we have no truly scientific idea of how consciousness emerges from material structures such as those of human beings. We can somehow associate electric signal to emotions but in doing so we look at emotions as if they were objects. The intrinsic "true" uniqueness of human consciousness (and most likely animal consciousness as well) escapes our scientific understanding.

However, even given the limitations of scientific theories, the scientific method still holds:

- A theory that is not in agreement with observations is not a theory at all;

- A theory that does not allow one to forecast the behavior of the object under study is basically useless;
- A theory that cannot be falsified by any observation is an intellectual exercise void of content;
- A theory formed with vague concepts is not a theory at all, perhaps only the expression of an ideology.

Note that the scientific method does not necessarily require that theories be expressed in mathematical formalism. Modern physical science is generally (but not always) embodied in mathematical formalism. Botanists might have knowledge of physical phenomena expressed in plain English. They might be able to make useful predictions regarding the growth of trees or other plants but be unable to express their knowledge in mathematical models. Their statements can nevertheless be empirically tested.

In our attempt to understand the impact of money on real economies using the scientific method, we adopt the view that economies are complex systems. The theory of money needs to be formulated as a global theory based on observations and *linked to economic theory*. Money is not designed on the basis of theoretical principles, with a clear definition of objectives; money is a practical tool that evolves in response to economic and social forces. A theory of money is an attempt to understand how complex economic and social phenomena interact and evolve.

Generally speaking, in textbooks on economics and banking, money is defined through its three main functions: (1) as a means of exchange, (2) as a store of value, and (3) as a unit of account (i.e. measure of value). While there is some agreement on the functions of money, what underlies these functions is subject to debate. Two main views have been advanced:

- Money is *something with intrinsic value*, such as gold or silver;
- Money does not necessarily have intrinsic value: it can be either *fiat money*, that is, a token that acquires value by the decree of an issuing authority, or it can be *credit* or, more exactly, a certificate of purchasing power.

In addition, there is now (almost) a continuum of highly liquid financial assets that can be defined as money or near-money. What we define as money is partially conventional. Any theory of money depends on one's definition of money. We adopt three different perspectives on money:

- a historical perspective;
- an engineering and institutional perspective;

- a philosophical perspective.

The historical perspective looks at the development of money throughout history. In particular, it tries to understand the evolution of money and its role in societies, not out of pure curiosity but because it informs us on the many forms and uses of money as a response to social needs and objectives. The engineering and institutional perspective looks at money as an attempt to create the tools and institutions needed to facilitate exchange and, in our modern economies, the planning of production and the distribution of the fruit thereof. The philosophical perspective looks at money as a universal exchange medium of human societies (assuming this to be the case). We will try to highlight, wherever possible, the three perspectives. In particular, we will look at the current financial system from an engineering perspective, exploring how it functions today and how the system might change to meet the challenges of tomorrow.

In our analysis of the theory of money, we propose a framework based on the following five questions:

- What is money?
- How is money generated?
- How is money distributed?
- How does money acquire value and how does that value change?
- How does money impact the economy and society?

Chapter 1 presents the basic concepts of money. We start by analyzing what characteristics of a society require the use of money. We will see that money is a requisite in human societies characterized by ownership and exchange, with the concept of ownership predating that of money. Money is the medium that allows a society wherein individuals (be they persons or entities such as firms) own and freely exchange goods and services and plan production. In the absence of money, economic life is organized by command. This consideration leads to the study of free markets but we will see how, inside free markets, there are entities that are in themselves command economies. The coexistence of free markets and command economies seems to be an essential characteristic of human economic organization.

Chapter 2 discusses a number of basic topics in formulating a theory of money. After arguing that we need a theory of money given the ability of modern economies to create and distribute money, the chapter discusses Operationalism as a key methodological principle. We next present the notion of stock-flow consistency and how money theory integrates with macroeconomics. The chapter closes with our framework of analysis for the theories of money.

Chapter 3 opens with a discussion on the nature (or forms) of money, from barter

to the modern concepts of fiat money, credit money, bank money, and digital money. We will clarify in operational terms the many sources of potential confusion between the concepts of fiat and credit money, paying particular attention to the notion of credit money. The idea that money is something that gives its owner the ability to purchase goods and services will be contrasted with the idea that a bank account is credit because it can be converted into banknotes.

Chapter 4 presents early mathematical models of money. These early models were created with the intent of integrating money into general equilibrium models of the economy.

Chapter 5 discusses how money is generated with particular attention to its generation in the modern banking system. Our discussion includes a number of conceptual and practical issues in modern systems where money is considered a "creature of the state" but is primarily generated by commercial banks. The nature of money as credit raises a number of questions including the role of base money (i.e. coins and banknotes) and the fact that money might become such a poor store of value that near-money becomes the preferred store of value. The two frameworks that compete to explain money generation – exogenous versus endogenous – are discussed.

Chapter 6 discusses how money acquires value and how that value might change over time. We argue that the nominal value of money, that is, the purchasing power of a unit of money, is determined by the past history of the currency though it can be arbitrarily modified by the state at any moment as demonstrated by the introduction of the new French Franc (1960). In our discussion of how value changes over time, our focus is the problem of measuring inflation in a complex economy, critical for any theory of growth.

Chapter 7 discusses the distribution of money, which determines the allocation of resources and ultimately the creation of wealth and economic equality/inequality as well as the economic power that goes with it.

Lastly, Chapter 8 discusses the interaction of the dynamics of money with the dynamics of the real economy. Most of today's models of an economy do not include a banking system, the assumption being that markets clear instantaneously in a situation of equilibrium. But there are difficulties in reconciling an equilibrium model with a theory of money originated as credit. Based on the theoretical insight of the theory of money as a circular flow (Circuitism), we introduce stock-flow consistent models. Stock-flow consistent models include a banking system and, unlike general equilibrium models, are able to model financial instabilities. These models include all the essential features of an economy based on credit as well as the economic complexity that characterizes modern economies, including the emergence

of financial and/or economic instabilities.

Let’s then begin with Chapter 1.

Note

1 Even in the hard sciences, new theories are rarely enthusiastically accepted when first proposed. The German physicist and recipient of the 1918 Nobel Prize in Physics Max Planck (1858–1947) famously saic "Science advances one funeral at a time."

Reference

Quine, Willard van Orman. 1953. “Two Dogmas of Empiricism” in *From a Logical Point of View*. Cambridge, MA: Harvard University Press.

1
The Theory of Money

Basic concepts, part I

1.1 Introduction

This chapter lays the foundation for our discussion of money. It introduces some of the basic concepts of the theory of money, including its nature and function, the processes through which money is generated and distributed, and the interaction of the dynamics of money with the dynamics of the real economy. This chapter highlights the key themes of the book that can be summarized as follows.

1.1.1 On the nature, creation, and distribution of money

Money is essentially a token that represents the purchasing power of its owner. Money requires a pre-existing structure of ownership of goods, reinforced by institutions. Money does not create ownership but allows the transfer of ownership. The creation and allocation of money solve the engineering problem of making available to individuals and entities the tools necessary to exchange the ownership of goods and services and to store purchasing power for the deferred exchange of ownership.

Textbooks identify three main functions of money: (1) as a means of exchange, (2) as a means of storage, and (3) as a unit of account. For our purposes, we propose the following definition:

> *Money is a tool that allows implementing economic decisions freely and autonomously by transferring the ownership of things or granting the fruition of services.*

To the three main functions of money, we add the concepts of ownership and free exchange. Money is enabled by ownership and by institutions that protect ownership. Money is also enabled by the willingness of economic agents to transfer ownership to others in exchange for money. Ownership logically precedes money.

As Martin Shubik (2010) remarked, money implies trust. A person or an organization gives up the ownership of something in exchange for money under the assumption (trust) that the same money will be accepted in future transactions and will allow them to acquire ownership of other goods or services of similar value.

Having outlined the basic functions of money, we can now ask: What is the nature of money? There are two main historical views on the nature of money: (1) money as something with intrinsic value, for example precious metals such as silver or gold; and (2) money as something with no intrinsic value such as cowry shells, banknotes, or bank deposits that represent the purchasing power of individuals or entities such as firms. Shubik suggests that the backing of money with something that has intrinsic

value is one way to gain trust for the money. However, in advanced economies, money in the form of banknotes is no longer convertible into something with intrinsic value and now represents only a small percentage of the monetary mass (3–10%). Most money is now in the form of bank deposits. The forms of money continue to evolve over time and the present trend is clearly towards its dematerialization.

As observed in the Introductory Remarks, there are three perspectives on money: historical, engineering/institutional, and philosophical. The historical perspective seems to suggest that the current view of money as credit is a natural evolution of the practice of credit. Anthropological studies suggest that credit and debit were the first means of exchange though what concrete form they took varies. The moment it was realized that credit is transferable, the way was paved for introducing the general notion of money as credit.

Today, however, central banks are confronted with the engineering problem of deciding if and how to dematerialize money, eliminating banknotes and eventually implementing institutional changes. Introduction of digital currencies is one such major change. The engineering problems that such a change would involve would require a higher-level view of the question of money. In fact, governments and central banks would have to rethink the problem of how to distribute the means of payment to a large population within an economy that, given present trends in automation, will likely leave a large fraction of the population without sufficient means to live a decent life. In republican Rome (123 BC), the solution to feeding the large population was to provide a monthly distribution of grain – the cheapest and most efficient foodstuff then available – at a set cost (6 and 1/3 asses per modius [1]). The program was financed by the reorganization of the taxation system in the rich Asian provinces (Aly 2017).

Money is by nature *hierarchical*, that is to say, money at a given level depends on money at a higher level. Dependence might take different forms: for example, money of a lower level might be convertible into money of a higher level or might be a package of credits such as a money market fund formed with short-term treasury debts. At the top of the hierarchy is money with the highest level of liquidity; at lower levels, money consists of progressively less liquid assets. Today, the hierarchy of money consists of coins and banknotes at the highest level, bank deposits at a lower level, and near-money assets such as money market funds at progressively lower levels. Bank deposits are presently the most common and widely used money for transactions.

In today's economies, the production of money either as banknotes or as deposits is a relatively simple, low-cost process while the allocation of money is a complex process, the key element being credit. New money created by minting coins or

printing banknotes, hierarchically superior, is produced by central banks, "sold" to banks, and can be obtained by the public only from the banking system. Coins and banknotes are permanent money. New money as created when a commercial bank extends a loan is not permanent money. Loans are created as deposits and extended to creditworthy agents (individuals or entities) and must (in time) be paid back; once paid back, their value is destroyed. Deposits can of course be transferred between different bank accounts totally or partially, but transfers do not create new money, they only transfer money. Near-money, such as money market funds, is a complex financially engineered asset sold in the market and used as storage of value. Though the generation of near-money is complex, its distribution is a straightforward sale.

Systems for the creation and allocation of money are not fixed but are in a constant state of flux. Future evolutions might include the abandonment of coins and banknotes mentioned above or the abandonment of the strict reliance on debt to allocate money, replacing the latter with other mechanisms such as the uniform distribution of money.

Clearly, designing the process of the creation and distribution of money is both an engineering problem and a problem of political economy. In fact the creation/ distribution of money impacts the distribution of wealth and the stability of societies such as our democracies. The Classical Greek philosopher Plato believed that to ensure social peace, the income of the highest paid in society should never amount to more than five times that of the lowest paid. We might disagree with the exact figure but the principle is clear.

1.1.2 On money and prices

The relationship between money and market prices is complex. The Quantity Theory of Money (QTM) posits a simple proportionality relationship between the quantity of money in circulation and the level of prices where the level of prices is, in practice, a weighted average of prices. Though intuition and experience suggest that, other things being equal, prices increase as the amount of money available increases, the QTM seems an oversimplification.

Economic theory posits that prices are determined by the intersection of supply and demand. But demand depends on the amount of money available for purchases. Modern economies – highly unequal in terms of income and wealth – are segmented in sub-economies that have only marginal mutual interactions. The notion of a single price level for the entire economy does not correspond to empirical reality. First, the choice of a price index is largely arbitrary; different formulas can be adopted. Second, assuming that we have chosen an index, each sub-economy has a different

price level, receives different inflows of money, and reacts differently to changes in the inflows of money. We suggest that the notion of an average relationship between the price levels and the quantity of money would better be replaced with a vector of different relationships between money and prices for each sub-economy.

1.1.3 On inflation, growth, and financial and economic crises

Modern economies are complex systems that output complex products and services. The market value of these products and services is only partially related to their physical characteristics. Markets attribute values to products in function of many factors, including the image of the product and its manufacturer. Given the level of innovation and the weak relationship between physical characteristics and market value, it is impossible to compute a unique true inflation rate.

Therefore, there can be no unique true distinction between real and nominal growth, as posited by mainstream growth theory and discussed in Barro and Sala-i-Martin (2003). Nominal growth is due to both monetary and physical factors, the latter including changes in quantities produced and in the complexity of products. In modern economies with a high rate of innovation even in staid sectors of the economy – be it goods or services – distinguishing real growth from nominal growth is somewhat arbitrary.

Different sub-economies might grow at different speeds. In particular, inflows of money might produce a rate of growth of financial markets that far outpaces the rate of growth of the economy, as we have seen over the period 2007–2016 with U.S. financial markets growing three times as fast as the real economy. These differences in growth rates within the same economy might produce the instabilities described by Hyman Minsky, resulting in market crashes and eventually economic crises (Minsky's theory of financial instability will be discussed in Chapter 8).

This chapter gives an overview of these themes; their development constitutes the rest of the book.

1.2 Can we do without money?

Let's begin our exploration of the nature and function of money in modern economies by considering a society without any form of money. Our objective is to ascertain what characteristics of human societies enable or require the use of money. In a nutshell, what characterize a society with money are a structure of ownership and the freedom to exchange goods and services.

To gain a better understanding of money it is useful to explore if and how societies can function without money. This implies identifying social structures where there is no (or a very limited) concept of ownership and where important economic decisions are implemented with tools other than monetary exchange. We can identify three basic types of societies without money:

- gift economies;
- centralized economies where economic decisions are made by central authorities and implemented via orders of some type;
- and, theoretically, post-scarcity societies where everything needed is freely available.

Consider first what anthropologists call gift economies. Gift economies, if they existed/exist, are primitive societies where goods are shared without barter or compensation. A well-known example, popularized by the BBC documentary *Tribe*, comes from the Anuta people, a tribe of 300 persons living in the Solomon Islands in Oceania. The Anuta have a social life based on the notion of sharing and gift-giving.

The reality behind gift economies is, however, much debated. The anthropologist Bronislaw Malinowsky (1922) studied the Trobriand people off the east coast of New Guinea and concluded that what was considered a gift economy was in reality a complex structure of power and politics. However, for the French anthropologist Marcel Mauss (1923), considered one of the founders of anthropology, the role of the exchange of gifts was to create bonds in human societies. Nephew of the philosopher and sociologist Emile Durkheim, Mauss formulated broad theories of gift-giving in primitive societies. He posited that a "gift" was not really a transfer of ownership as any gift made must be reciprocated, perhaps not under the same form.

Anyone who is familiar with life in European country or mountain villages has probably remarked that gifts are rarely welcome: receiving a gift binds one to reciprocate, possibly with a gift of a slightly larger value. In her study of traditional Japanese culture researched for the American Office of War, the anthropologist Ruth Benedict (1946) describes how, in traditional Japanese culture, people do not come

to the assistance of others to avoid binding them with the duty to reciprocate. Gifts and assistance might create unwelcome bonds.

But what if there is no money in a modern economy where exchange takes place? One way to eliminate money is to eliminate ownership, centralizing it in the hands of a group of people such as a government or an army. This ruling group might then have products or services produced and distributed without the use of money, by centralized dictate, such as deliver x amount of y to z. Economies of this type are called planned (or command) economies. An example of a command economy comes from Ancient Egypt, under the Pharaohs who, in theory at least, owned all the land and in whose granaries and treasuries surplus produce was stored. Employees in the thousands working in noble or royal households were guaranteed sustenance from these surpluses. Gordon Childe (1982 [1942]) gives us a precise idea of how the labor of each of the thousand workers who were building King Seti's temple in the second millennium BC was valued in terms of commodities: "'4 lb. bread, 2 bundles of vegetables and a roast of meat daily, and a clean linen garment twice a month'!" (p. 131).

In modern times the classical examples of command economies are the Soviet Union, China, and, on a smaller scale, Cuba. While it is not our objective herein to analyze the characteristics and the history of these command economies, let's remark that the Soviet Union, China, and Cuba were not economies without money. There was a very high level of central planning as regards, in particular, agriculture and industry but people were allowed, albeit with constraints and limitations, to use money to purchase goods and services.

It is often remarked that a planned society cannot gather the information necessary to its functioning and eliminates the freedom of the individual to decide the course of action he or she prefers and that this freedom is accorded by ownership and money. But this is not necessarily the case. We can envisage planned societies where individuals still have considerable freedom of action, as we will see. It is true, however, that coordinating the supply of goods and services to a large segmented population such as that of modern developed economies would pose an enormous challenge in terms of information gathering.

1.2.1 Religious organizations

One example of large organizations whose *internal* structure resembles that of a command economy is religious organizations. Consider, for example, the Catholic Church, an organization almost 2,000 years old. True, the Catholic Church like others

has amassed great wealth, but the function of the thousands of persons inside the organization is determined by command. Consider that the ordained clergy alone counted 465,595 persons (or employees) in 2014.[2] When an individual enters into the service of the Church (takes orders), his or her life is regulated by rules; remuneration is for the most part not in money but does include food, clothing, and lodging. Rituals such as baptism and communion, and interaction with the faithful such as visiting the sick, are activities performed by the Church's personnel but not directly in exchange for money.

Religious organizations of course need money to buy goods and services from the outside world. This money was generally provided by donations or tithes imposed by rules that forced members of a society to give a contribution to the cult. In the past, in many European countries persons had to give one tenth of their "income" (the *decima*) to a Christian cult. Still now, in countries such as Germany, taxpayers designate a church or a cause as the beneficiary of a small fraction of their income taxes. This, however, does not change the fact that a large and complex organization can work and prosper without the need to exchange money inside its own structure.

Anticipating a theme that will be developed in the next section, religious organizations are examples of command societies embedded in market economies. Ultimately it might be that in the modern world it is more typical to find a mix of the two types of economies, free-market and command.

1.2.2 Firms: the invisible, visible, and vanishing hand

Another example of how large organizations can function internally without the need to exchange money comes from the business world. In the popular press and in the perception of most people there is a dichotomy between free-market capitalist economies and centrally planned Soviet-style economies. This distinction is primarily political and ideological. Academic research has discussed a more common implementation of planned economies: large integrated firms.

In his *The Wealth of Nations*, Adam Smith described a hypothetical "pin factory" where specialization and division of labor allow a firm to attain a high level of productive efficiency. However, despite the efficiency of the pin factory, Smith describes how free markets optimally coordinate economic activities as if they were guided by an invisible hand. Classical and neoclassical economists successively developed the notion of free markets as self-organizing and self-regulating optimal allocators of resources.

It was the English-born economist Roland Coase who first described the

opposition between the concept of free markets and the notion of "firms." In 1937, the 27-year-old Coase published his paper "The Nature of the Firm" where he put forward the notion that there are two organizations of economic activity: (1) competitive markets coordinated by price and (2) firms coordinated by the conscious activity of the entrepreneur. Coase (1937) was trying to answer the question of why we find the "islands of conscious power" that the economist Sir Denis Holme Robertson had observed in his study of firms operating in free markets. Robertson (1923) wrote that we find "islands of conscious power in this ocean of unconscious co-operation" (p. 85). Coase's answer was that firms exist because in free markets there are transaction costs such that the internal organization of a firm is more efficient than the market alone.

Initially ignored, the work of Coase was later considered a milestone in economic theory and the 81-year-old Coase was awarded the Nobel Memorial Prize in Economic Sciences in 1991. Coase's work was expanded theoretically by one of his students, Oliver Williamson, who studied the decision-making process of markets and firms. Williamson was in turn awarded the Nobel Memorial Prize in Economic Sciences in 2009.

With the development of large corporations, the economic historian Alfred D. Chandler (1977) described the emergence of a new type of corporation, integrated and with layers of middle management. Chandler concluded that the history of American capitalism runs contrary to the increasing specialization and division of labor foreseen by Smith. He argued that large vertically integrated corporations have replaced large areas of free markets with well-organized command structures. Smith's invisible hand is thus replaced by the visible hand of layers of middle managers who run large vertically integrated corporations.

More recently, Richard Langlois (2002) observed that Chandler described the situation prevailing in a specific historical period while at the turn of the 21st century vertical integration was being replaced by other forms of organization.

Without going into detail on how firms are organized, it is clear that a large fraction of economic activity is not performed by free markets coordinated by price where money is the key transmitter of free, autonomous economic decisions, but by firms with internal, decentralized decision-making that does not use money to implement decisions. In general, such decentralized organizations work by assigning budgets which typically include a notion of value but rarely include actual transfers of money. When they do so, it is more for fiscal than for operational purposes.

The above examples show that it is possible to manage large, complex organizations with thousands of employees and (de)centralized structures that exchange among themselves, using an internal notion of value but without the use of

money: activities are coordinated by internal service orders and (possibly) the autonomous decision-making of individuals.

Planned (or command) economies are typically associated with the Prussian-born (political) economist Karl Marx and communism. It is widely believed that in planned economies individuals have little to no freedom. Planned economies came to be identified with oppressive regimes while market economies were identified with freedom, democracy, and prosperity. The 20th-century economists-philosophers Fredrick von Hayek and Ludwig von Mises were strong critics of centrally planned economies. Their criticism hinged on the belief that a centrally planned system destroys freedom and cannot gather and process the information needed to plan an advanced economy. However, as we have seen, planned sub-economies are common even within capitalist societies.

1.2.3 Post-scarcity societies

A modern, futuristic approach to an economy without money is to envisage a society where technology has reached the stage where all needed goods and services are freely available to everyone, with money relegated to irrelevance. An integral part of Karl Marx's notion of social progress included a generous post-scarcity society where all needs would be satisfied without recourse to money. A similar post-scarcity society was envisioned in the *Star Trek* series and films. In the *Star Trek* world, "replicators" produce anything needed. The post-scarcity, moneyless economic system in the *Star Trek* universe was discussed at New York Comic Con[3] in October 2015. Among the panelists was Paul Krugman, recipient of the 2008 Nobel Memorial Prize in Economic Sciences. Krugman remarked that a post-scarcity society in which objects are freely available would cover only 30% of Americans' spending; the remaining 70% is spent on services. Krugman observed that if *Star Trek* replicators had also to make available services, then the difference between servitors and slaves would be blurred. Krugman's remark implies that the replication of objects can be thought of as purely mechanistic while services are associated with human-like conscious abilities.

Besides the technical challenge of a post-scarcity society – a challenge possibly unsolvable – the cultural and psychological changes called for by a post-scarcity

society would be a critical issue. Most products used in our modern economies have a symbolic value insofar as they reflect the wealth of their owner. A post-scarcity society would eliminate social differences due to income.

Box 1.1 Will Robots Replace Humans?

There is currently much debate as to just how far the robotization of the workplace will go, though we are still far from being able to replace humans with robots in high-level intellectual tasks. Whether or not it will be possible to build a silicon brain with abilities similar to a human brain is a question that, for the moment, has no answer.

We have an idea of how a brain works only at the level of neurons and electro-chemical exchanges between neurons. A brain formed by some 100 billion neurons (the most recent estimate published in Azevedo et al. [2009] put the count at 86 billion neurons) is a complex system.

In addition we have no idea if the *workings* of a human brain can be explained as an assembly of neurons. While Ray Kurzweil (2013), the director of engineering at Google, in his book *How to Create a Mind*, argued that it will be possible to create machines more intelligent than humans, there is no consensus that this is effectively possible. Nor is there proof that high-level intelligence can be explained in computational terms. In addition, we still have no knowledge of how a conscious mind is related to a physical brain.

Surely the brain is a physical object whose functioning can in principle be explained by physical laws but there is no guarantee that the brain is a computable system that can be simulated with a computational device. The brain, as an assembly of elementary particles, might simply be too complex to compute by any computer.

The challenges in creating a moneyless post-scarcity society can be seen in the present diffusion of automation. In fact, while post-scarcity societies have obviously not yet come into existence and it is unlikely we will see them in the near future, what we are seeing is the progressive introduction of automation – robots as well as other industrial automation devices in the workplace and artificial intelligence in the office – to replace humans. The loss of jobs could be enormous. Consider what happened when agriculture was mechanized. Patricia Daly (1981) estimates that in 1870, 50% of the U.S. population was employed in agriculture. Following the mechanization of agriculture during the period 1870–1981, Daly estimates that only 4.5% of the population worked in agriculture by 1981. Today that figure is approximately 1.5%. Figures for the rest of the developed world are not significantly different.

Our social values and structure lead us to turn to automation to increase profits, not

to reduce labor and distribute income. The present trend towards automation entails a contradiction: in return for their work, humans receive wages that allow them to consume products and services, machines do not. A highly automated economy organized to maximize output with the minimum cost will encounter insufficient demand as increasingly large fractions of the population will have neither jobs nor income. Implementing a peaceful highly automated society will require major cultural change.

1.3 Money, markets, and value

Let's now discuss how money as credit might become a source of economic instability and how the unequal distribution of money affects pricing in market economies as it segments the economy into sub-economies that have only marginal interactions and experience different growth rates.

1.3.1 Economic and financial instability

Following Keynes, Minsky (1986) provided the insight that a capitalist economy based on credit is essentially unstable, and this is because credit is prone to create what is called an autocatalytic process where an increase in credit fuels an additional increase in credit in what he likens to a Ponzi scheme.[4] Minsky argued that the main source of instability comes from the fact that credit can be granted based not only on the ability to repay interest with real gains but also on an anticipated future increase in the value of assets. (The instability thus produced is discussed in Chapter 8.)

Box 1.2 The Creation and Allocation of Money in Modern Economies

The problem of money in modern economies is the engineering problem of how to create and allocate means of payment to a large number of individuals and entities. There are several forms of money in modern economies: coins, banknotes, bank reserves, bank deposits, plus near-money and alternative and digital currencies. No current form of money has intrinsic value; all forms of money circulating today can be thought of as tokens or certificates that establish the purchasing power of its holder. Coins and banknotes are anonymous while the owner of a bank deposit is traceable. In the hierarchy of money, coins and banknotes occupy the highest level as they do not depend on other types of money. Reserves and bank deposits are convertible into banknotes; in principle they are credits/debts that depend on banknotes.

Reserves are money exchanged between commercial banks or between commercial banks and their respective central bank; in the latter case, they can be thought of as accounts held by banks with their respective central bank. In

aggregate, central banks are the exclusive creators of reserves. Reserves are created when central banks either buy assets from banks (or from non-banks using a commercial bank as an intermediary) or (occasionally) lend reserves to banks for (emergency) purposes. When funds are transferred from Bank A to Bank B, reserves are transferred from Bank A to Bank B; the total amount of reserves remains unchanged, only their position in the central bank has changed. When a bank creates a new deposit as a loan–deposit pair, no reserve is created or transferred.

Coins are minted by governments or central banks and distributed to banks in exchange for banknotes or bank deposits while banknotes printed by central banks are given to banks in exchange for reserves. Commercial banks in turn sell the banknotes to their clients, debiting their accounts. Both coins and banknotes are fiat money, that is, legal tender, thus both must be accepted in payment of goods and services including labor.

In many developed economies, including the U.S., the U.K., and the European Union, direct financing of the state is not allowed and therefore the state cannot distribute banknotes to the public in exchange for services. The public can obtain banknotes only by buying them from commercial banks. However, in some countries such as India, central banks might lend funds directly to their governments who then use them to pay for goods and services, including labor. In this case, individuals and firms might obtain banknotes by either buying them from banks or from the government in exchange for services.

The use of banknotes is in decline. In the U.K., as of year-end 2013, banknotes represented only 3% of the country's money stock; the figure for the U.S. is under 10% but half of this is held abroad. The vast majority of money is formed by bank deposits, which, being issued by banks, are not legal tender. However, deposits are universally accepted for payment, for example via bank transfers. In most countries, deposits are, in principle, convertible into banknotes. It is possible that sometime in the future banknotes will cease to exist; deposits will remain as records of the capacity of the holder of a deposit to purchase goods and services.

Banks create loans and deposits contemporaneously. Contrary to what is widely believed, banks do not need deposits to make loans and therefore do not need deposits to create money. Conceptually, we must separate the generation of deposits, which ultimately consists in banks writing numbers in their clients' records, and the allocation of newly created deposits to clients who take out loans. Other rules of allocation could be applied, for example the uniform distribution of deposits to citizens as a basic national income now in an

experimental phase in Finland and under study in the U.S. state of Hawaii.

Money is therefore temporary money as loans must be repaid and money destroyed. There is one important exception: money can be created/destroyed by central banks when they buy assets from or sell assets to banks in a process called open market operations. In the aftermath of the 2008 financial crisis, a significant fraction of bank deposits were created with money generated by the central banks in the process called quantitative easing (QE). In QE, a central bank purchases financial assets from non-bank organizations.

We can restate the problem in terms of money: the instability of a society based on credit is equivalent to the instability of a society where money is created by commercial banks as they accord loans. (See Box 1.2 for an overview of the money generation process in modern economies.) Instability is due to the economic tension created when the growth rate of the value of corporate shares (i.e. total market capitalization) is much higher than the growth rate of the real economy, thereby signaling a bubble in stock prices. This asset inflation is produced by an excess of money that flows into financial markets. Didier Sornette, a physicist-turned-economist, studied the mathematical characteristics of a stock market bubble with his colleague Peter Cauwels at the Zurich Polytechnic Institute ETH. According to thei analysis (Cauwels and Sornette 2012), financial bubbles are characterized by the super-exponential growth in the value of financial assets while the economy grows only exponentially.

Let's now expand Minsky's analysis in two directions. First, at the time Minsky made his analysis, credit was primarily extended to firms for financing their working capital needs as well as new investments. Today, firms are no longer the main borrowers. Adair Turner (2013) remarks that in the U.K only an estimated 15% of the total amount of credit issued goes to firms; most credit issued goes to households to finance mortgages. Much of this credit does not finance the building of new homes but the purchase of existing homes, creating asset inflation. This pattern can be seen in most advanced economies including the U.S. (for details on just who gets the borrowed money in the U.S., see Chapter 8). This notion can be generalized. Richard Werner (1992, 1997) introduced the notion that credit must be split into two parts: one part that finances transactions that impact the GDP, the other part that impacts asset prices. He applied his theory to the Japanese economy, showing how this split explains the reduced velocity of the circulation of money found not only in Japan but in most advanced economies. Recall that the velocity of the circulation of money is the stock of money divided by the GDP. In addition, the programs of asset purchases

implemented by central banks after the 2007–2008 financial crisis and subsequent Great Recession have created huge inflows of money into financial markets.

Second, we suggest that while Minsky considers the whole economy and the entire asset market, the analysis should be made at the level of subsystems. Why? First observe that high corporate profits and large gains in financial markets should in principle create demand for goods and services that would in turn create inflation as currently measured (see Chapter 6 for a discussion of inflation). As inflation is closely watched, governments and central banks would intervene to, as they say, "take away the punch bowl." But this is not what happened in recent financial and economic crises. In the U.S., for example, after the crash of 1987, inflation fell and remained low, on average below 3% until 2016. We have therefore to explain how, in the aftermath of the market crash of 2000 and that of 2008, low inflation coexisted with economic growth in line with historical averages but was accompanied by a steep rise in asset values. Exhibit 1.1 illustrates the growth of U.S. GDP and of the S&P 500 index for the period January 1987 to end-February 2016; Exhibit 1.2 illustrates the inflation rate for the same period.

We hypothesize that, in modern complex economies, economic and financial bubbles are explained by the complexity occurring in products, production, and economic relationships coupled with money generation. The fact that money generation can produce simultaneously an asset bubble, some economic growth, and no measured inflation is what makes crises difficult to predict as asset bubbles are sustained by several factors contemporaneously.

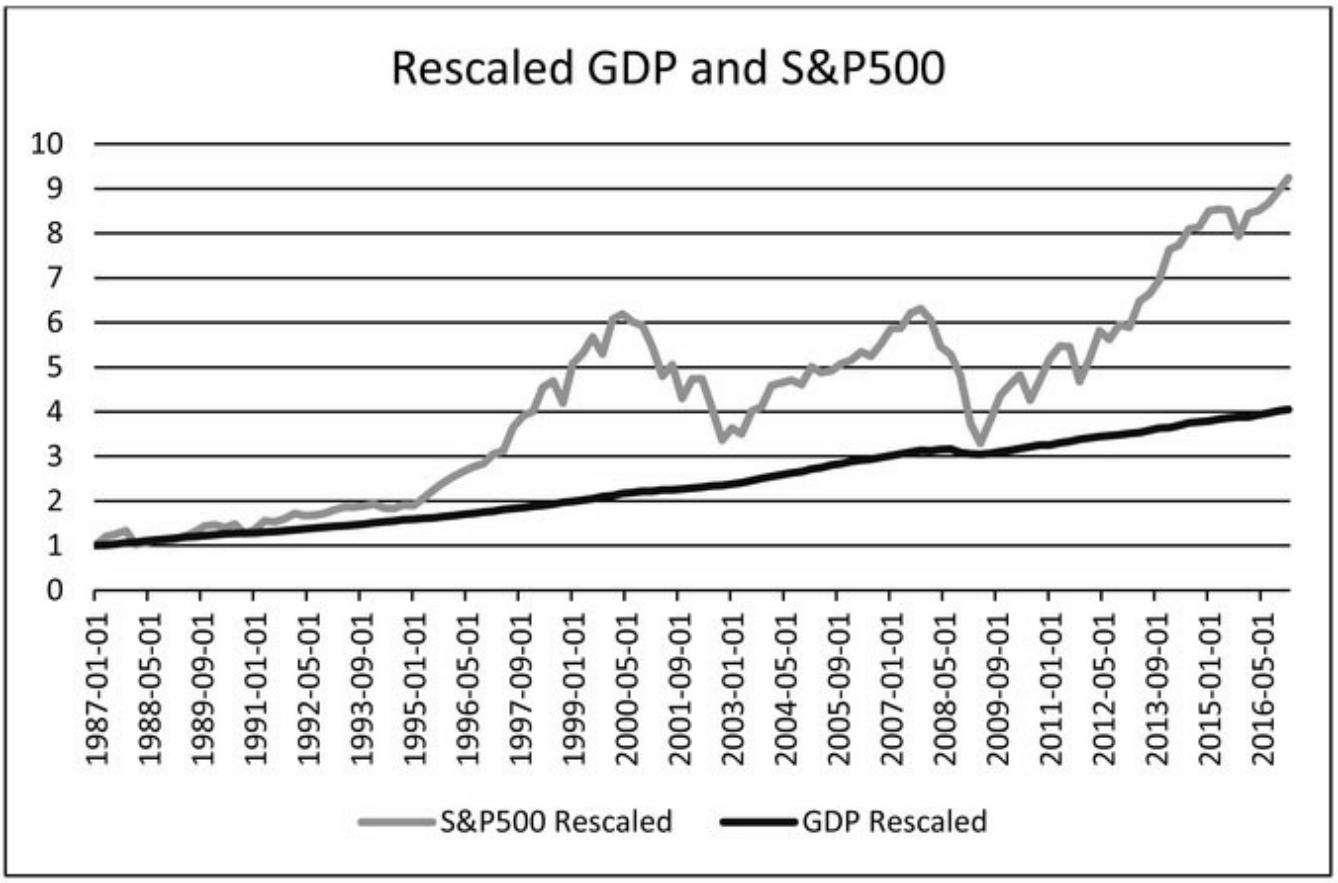

EXHIBIT 1.1 The growth of U.S. GDP and the S&P500 index for the period 1987–2016, rescaled to start at 1. Constructed by the author using data obtained from the Federal Reserve Bank of Saint Louis (FRED).

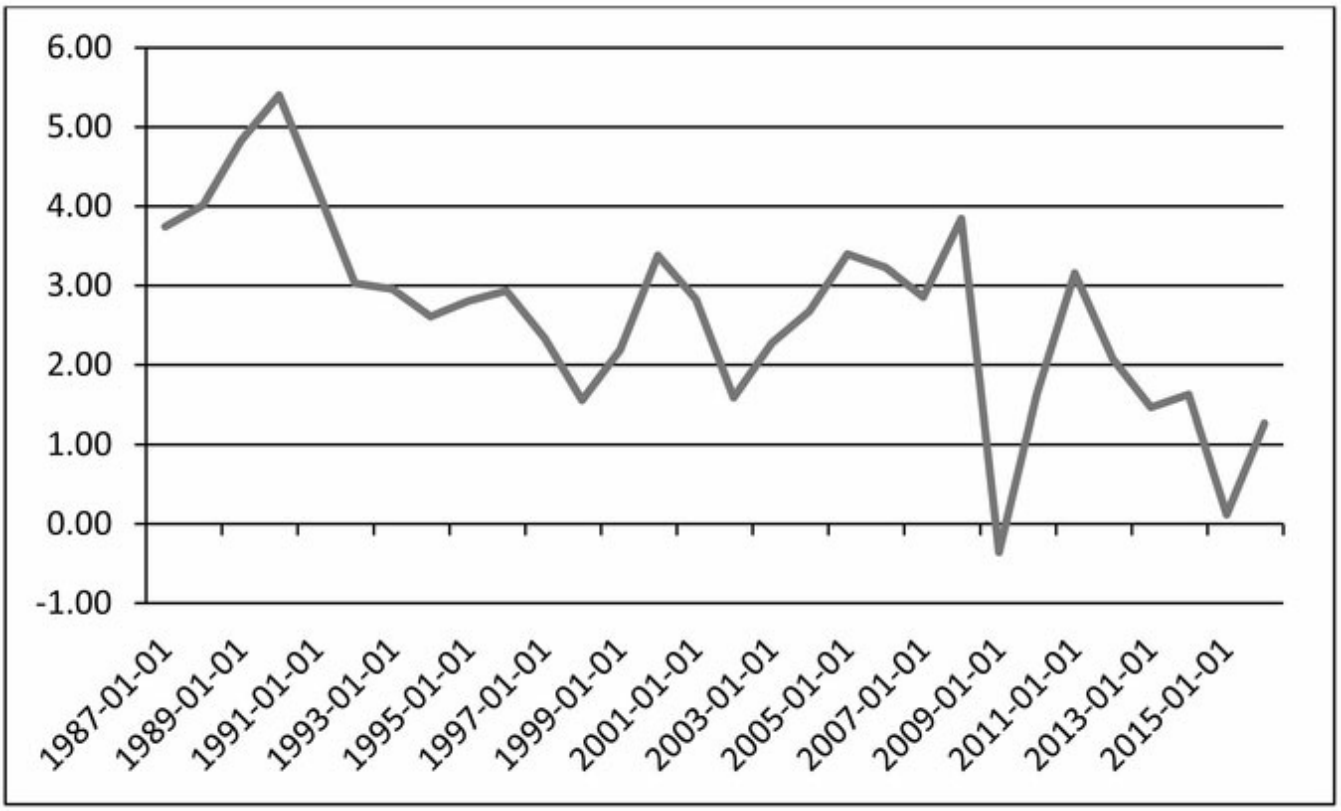

EXHIBIT 1.2 The U.S. inflation rate for the period 1987–2016. Constructed by the author using data obtained

from the Federal Reserve Bank of Saint Louis (FRED).

Our analysis hinges essentially on three points, all related to how we aggregate in a complex system. In particular: (1) there is no way to measure inflation uniquely for an entire modern economy, (2) money is distributed unequally, and (3) modern economies are segmented into sub-economies that grow at different rates. These three points are intrinsically related. Taken together, they explain how we might witness rapidly rising asset prices, some economic growth, and little inflation. Let's look briefly at each of these; a more detailed discussion can be found in Chapter 8, where we discuss the impact of money on economies.

First, inflation. The concepts of price indexing and inflation are not uniquely defined. We can define an infinite number of inflation indices and obtain different results for each index. What is generally measured by the usual consumer price indices applies reasonably well only to the most stable portion of the product space. There are entire sectors of the economy – the very innovative or the luxury sectors for example – where for one reason or another it is simply not possible to measure inflation. More generally, inflation should be replaced by a vector of inflation indicators.

As for the unequal distribution of money, there are two main sources of inequality. First, the division of labor between workers and capitalists defines the economic power structure and is responsible for how money is apportioned to wages and profits. Second, most new money is originated with bank loans, and loans are granted to creditworthy individuals or firms. With the low interest rates that have prevailed since the 2008 financial crash, most loans taken out by firms are now being used to buy back stocks. This is in itself an autocatalytic process where the growing price of assets creates ever-greater collateral and produces ever-growing quantities of money that flow into financial markets.

Third and lastly, the way free markets price products depends on how money flows into different sectors of the economy. If we have large inflows in sectors where inflation cannot be measured, we will experience nominal growth indistinguishable from real growth.

We can now complete our analysis. In the past three decades and more recently with central bank policies following the 2007–2008 financial crisis, large sums of money have flowed into sectors where inflation cannot be measured. The economy grew but only in those sectors where it produced a nominal price increase with little or no inflation. Some sectors of the economy experienced a nominal growth at a rate comparable to that of the financial markets. As the U.S. stock market valuation in 2007 was largely determined by inflows and leverage, the subprime crisis triggered

an economic crisis. Our analysis will be developed in Chapter 8.

1.3.2 The Quantity Theory of Money

Let's now see how our analysis differs from the Quantity Theory of Money (QTM) which states that, other things being equal, the price level is proportional to, or at least is a growing function of, the quantity of money. The QTM has its origins in the late Renaissance when the Prussian-born mathematician Nicolaus Copernicus and others noted the price increases following the inflows of gold and silver from the New World. In recent times, its proponents include Irving Fisher, Ludwig von Mises and Milton Friedman. We suggest that in modern complex economies, the notion of a market-wide price level is not meaningful; the QTM cannot apply to the entire economy. But we can recover some form of "local" QTM.

A market economy is an economy where individuals, be they physical individuals or other, own goods and freely sell their goods and services, typically in exchange for money; payment in kind is rare. So how does money work in a market economy? In a nutshell:

> *A market economy determines prices in function of preferences, the amount of money available, and of its distribution. Prices are not solely determined by the abstract preference ordering of individuals but by the amount of money individuals have at their disposal, that is, by the amount of money available to the various players in a market economy.*

The key differences of the above statement with regard to the QTM are (1) the recognition that money is unequally distributed and (2) the need to specify operationally how variables are measured. The QTM assumes that there is a relationship between the level of prices and the quantity of money available without considering the structure of wealth and income. But in the real world there is no unique well-defined level of prices applicable to the entire market and there is no uniform flow of money into the economy. Money flows selectively to some sectors and contributes to producing different growth rates in different sectors.

1.3.3 Capitalists, wage earners, and money

Market economies are typically associated with the free private enterprise system whose activity is the result of private initiative, without central control if not in the form of regulations. The free enterprise system, in turn, is associated with capitalism insofar as enterprises are owned by capitalists. There is a fundamental distinction

between capitalists who earn profits and employees who earn wages. In modern times, capitalists are either very wealthy individuals or professionally managed investment funds which are in turn owned by wealthy individuals.[6] Both are profit-seekers. The concentration of capital in the hands of a minority of persons or investment management firms in practice gives separate roles to capitalists and employees, thereby shaping the dynamics of market economies. Capitalists own the means of production and invest capital in a portfolio of enterprises with the objective of increasing their stock of capital; employees perform the needed labor and are paid wages (and perhaps a bonus) in exchange for their labor. We will call capitalists the ensemble of individuals and funds that own firms and seek profit.

Box 1.3 Money and Capitalistic Economies

Based on the notions of ownership, exchange, and production, free-market economies need money to allow capitalists and wage earners to organize (eventually long) production processes and the ability to exchange (buy) goods and services. Capitalists can accumulate physical capital or financial assets such as stocks, but in principle they do not accumulate money as money is borrowed and must be returned. Capitalists can accumulate money in aggregate, but only if the government runs a deficit monetized by the central bank or consumers take out loans. In fact, as we will see in detail later, all money comes from two sources: commercial banks when clients take out loans or central banks when they buy assets. Money created by commercial banks is destroyed when loans are repaid; money therefore cannot be accumulated. Typically capitalists accumulate near-money or other portfolios of liquid assets.

The division of the economy into capitalists and workers is a fundamental aspect of the power structure of modern advanced economies. It is this power structure that determines the apportioning of money, and therefore of economic output, between workers and capitalists. In a recent quarterly letter to investors, Jeremy Grantham of the Boston-based asset management firm Grantham, Mayo, Van Otterloo & Co (GMO) (2017) sought to explain why the price-to-earnings ratio (a measure of the valuation of a firm) of the 500 listed U.S. firms comprising the widely referenced S&P 500 index is now 65–70% higher than during the period 1935–1995. Grantham noted that compared to the pre-1997 period, the margins of S&P 500 firms have risen

by about 30% and identified six causes for this rise in profits: (1) the increased value of brands due to globalization, (2) increased corporate power over the past 40 years, (3) increased corporate wealth which has been used to influence policy, (4) a decrease in capital spending as a percentage of the GDP, (5) increased monopoly power for U.S. corporations, and (6) lower interest rates since 1997 together with higher leverage. High profits, a shrinking number of new investment opportunities, and the increase in money stock attracted growing flows of money to the same stocks, producing a price increase in excess of what is justified by corporate profits.

1.3.4 Are free markets really self-regulating?

Money based on credit might, as mentioned, add instability to an economy. The U.S. subprime mortgage crisis that triggered the recent global financial crisis is a good illustration of this process. What we will discuss in this chapter (and in more depth in Chapter 8) is how the distribution of money creates pockets of differentiated behavior in the product space and how, in each of these spaces, money affects the price level.

In classical and neoclassical economics, markets are endowed with almost magical powers: they are self-regulating and, if left undisturbed, tend to a stable economic optimum. A financial or economic crisis is therefore considered an exogenous event and excluded from their considerations. To illustrate the state of the "science," recall how, in the aftermath of the 2007–2008 market crash, the Queen of England intervened at a briefing by economists at the London School of Economics. Having reportedly lost some £25 million herself,[7] Queen Elizabeth II asked the economists why they hadn't seen the "awful" financial crisis coming. The committee of embarrassed professors hosting the Queen asked for time to prepare a written response. The answer arrived a few days later: "In summary, Your Majesty, the failure to foresee the timing, extent and severity of the crisis and to head it off, while it had many causes, was principally a failure of the collective imagination of many bright people, both in this country and internationally, to understand the risks to the system as a whole."[8] In essence, economists had not predicted the crisis because endogenous crises are neither imaginable nor foreseeable in the equilibrium framework of mainstream economics.

1.3.5 Local and global interactions and their role in determining prices

This and the following subsections discuss how money affects how free markets determine prices and highlight some fundamental differences between financial markets and markets for goods and services. Economists posit that price is determined by the intersection of the supply and demand curves. That is, it is posited that there are two monotonic functions of price, a decreasing function that represents the quantity that is demanded in function of price and an increasing function that represents the quantity that is offered in function of price. The actual market price is the price where demand matches the offer.

Economists further posit that economic agents order products and services in terms of their preferences. Under a number of mathematical conditions, the ordering of preferences can be expressed through a utility function that assigns a number, the utility, to each product and service. Demand is a function of the ordering of preferences. Is this a faithful representation of reality?

Two considerations need to be made. The first is that, in modern economies, preferences are only partially linked to the physical characteristics of products and services. Exogenous factors such as advertising, herding, or the symbolic value of goods and services might change the ordering of preferences. This decouples the economic value assigned to products and services from purely physical characteristics. Actually it might happen that the relationship between supply and demand is reversed as higher prices have more symbolic value. Consider, for example, Chivas Regal whiskey or Louis Vuitton handbags. In many cases production is limited by design to obtain a higher price. This is typical of "hot" electronic products such as video games or smartphones, typically before year-end holiday gift-giving. The physical growth of output, in terms of both complexity and quantity, is only weakly related to the economic growth.

The second consideration is that preferences are primarily determined by local comparisons. The notion that, in a free market, prices are determined by market-wide comparisons does not make sense in modern economies. Per Bak, Simon Nørrelykke and Martin Shubik (1999) created a multiagent model of price formation based on local interactions. They write:

> In equilibrium theory, all agents act simultaneously and globally. In reality, agents usually make decisions locally and sequentially. Suppose an agent has apples and wants oranges. He might have to sell his apples to another agent before he buys oranges from a third agent: hence money is needed for the transaction, supplying liquidity. It stores value between transactions.
>
> *(p. 2528)*

Locality in determining prices is different in financial markets and in markets for goods and services. In principle, financial markets are priced more uniformly across wide geographical areas than markets for physical goods. Though, like consumer

markets, they might be influenced by fads and herding (consider the late 1990s dot. com bubble), overall, financial markets are more coherent than markets for products and services. The reason for this coherence is that the price of financial assets depends on basic discount factors to be applied to the cash flows that characterize the asset. This is important: it allows us to reasonably define a concept of asset inflation and understand how financial markets respond to inflows of money while it is basically impossible to do the same in markets for real goods and services.

1.3.6 Asset pricing in financial markets

Financial assets do not have the many different physical characteristics of goods such as consumer products. In fact, a financial asset is a contract that gives its owner the right to receive a future stream of cash flows. This definition applies to any financial asset – bonds, stocks, or derivatives. Future cash flows are uncertain and are therefore represented probabilistically as random variables. Different financial assets are distinguished only by different streams of cash flows.[9]

Box 1.4 The Efficient Market Hypothesis

Eugene Fama (1970) introduced the Efficient Market Hypothesis (EMH) which states that prices fully reflect all available information. But Fama himself soon realized that this statement is vague and that testing the EMH requires a model of asset pricing. Fama and his followers subsequently reformulated the EMH, stating that markets are efficient if stocks are fairly priced, that is, if a stock's market price is roughly equal to the stock's intrinsic (fundamental) value. The definition of the EMH has evolved since its original formulation. Currently the EMH is defined in terms of predictability of asset prices; in the current literature, the EMH coincides with absence of arbitrage.

The EMH has become a cornerstone of finance theory, used to determine the fundamental value of a financial asset. Our concern here is not in discussing the EMH but with how markets price financial assets. How can financial assets be priced by demand and supply and still have a theoretical, intrinsic value perhaps far from its market price?

In the case of financial assets, supply and demand determine the discount factors used to determine the present value of future cash flows. In the absence of arbitrage, asset prices can always be computed as the sum of the present

discounted expected value of future cash flows. An efficient market is a market where the discount factors are those of a market in equilibrium, where the demand and supply of assets are also in equilibrium. Markets are efficient if market prices are aligned with equilibrium prices.

Pilkington (2014) recast the debate on EMH firmly in the framework of Wicksell's natural rates of interest, that is, those rates where savings equal the demand for investments. If there is no net inflow or outflow of money in search of investment, we can assume that investors rebalance their portfolios for technical reasons, responding to, for example, news. It is possible (but not certain) that trading, even for technical reasons, leads to a generalized increase in prices. However, if there is a net inflow of money into financial markets, we can assume that, in general, prices will increase.

Asset pricing theory states that, in absence of arbitrage, the market price of any financial asset is the sum of the expected values of all future cash flows discounted by an appropriate discount factor which reflects both the risk-free rate and the risk of the asset. Note that from a scientific point of view this notion is weak: a possibly infinite future stream of cash flows is not observable. Future cash flows are not predicted by finance theory. Therefore the theory of asset pricing is not an operational theory. However, in mathematical finance the notion of asset prices as the sum of discounted expected values of future cash flows applies to any price in the absence of arbitrage.

How does this relate to supply and demand? Modern stock exchanges – often automated auctions – implement the principle of supply and demand. They keep books of buy/sell orders and try to match the two. If selling prices are too high with respect to buy orders, the exchange waits until selling prices are reduced and/ or buying prices are increased.

The discount factors, as mentioned above, have two components: a risk-free discount factor and risk premia. The risk-free discount factor is common to the entire market; risk premia are idiosyncratic for each type of risk but still need to respect absence of arbitrage conditions. Clearly the pricing of financial assets is a complicated process that depends on central banks fixing basic discount rates as well as on market forces that determine market rates. However, arbitrage arguments suggest that financial markets should be priced according to uniform rules for the entire market. As asset prices are influenced by basic rates, we can expect financial markets to react uniformly to inflows/outflows of money.

The consequence is that it is easier to define an asset inflation index than a product

inflation index. In practice we can choose some index that represents capitalization as the global measure of the level of asset prices. Indices such as the S&P 500 or the Russell 2000 represent a possible measure of the level of the market. We can conclude that pricing in financial markets works by the general law of supply and demand and that financial markets respond in a rather uniform way to inflows of money. It is therefore possible to discuss a reasonable concept of asset inflation.

1.3.7 Pricing goods and services in competitive markets

Let's now discuss how competitive markets price goods and services. These markets can be more or less transparent in the sense that the price at which products and services are offered might be known to all market participants or might be available only to some market participants. Transparent markets tend to work by the rule of "one price," that is, the same thing has the same price in all markets.

There are, in practice, many deviations from the rule of one price. Department stores, for example, have pricing rules that might depend on the client, the day, the amount of time an item has been in the store, etc. Or one might find the same product offered at different prices in different stores. Different travel agencies might offer the same trip or the same hotel room at different prices, while the price on the Internet of airline seats and hotel rooms changes in real time as a function of supply and demand. The strategy of adapting prices in real time to changing demand is called dynamic pricing.

Consider a large modern economy. According to Eric Beinhocker (2007), in a modern market economy there are hundreds of millions of Stock Keeping Units (SKUs) whose function is to uniquely identify products. Assuming only 100,000,000 SKUs, there are 10,000,000,000,000,000, that is, 10,000 trillion relative prices. Markets are supposed to perform the computation – rather implausible – of correctly establishing tens of thousands of billions of relative price comparisons.

But this is of course not what happens. Market prices are simply the result of relative partial comparisons. There is no super-computation that finds the *one correct* price. There are millions, or perhaps hundreds of millions, of local and Internet comparisons between restricted sets of products. All these local comparisons generate the demand that, when matched with supply, ultimately produces the price.

1.3.8 Value and the segmentation of the economy

Value is market value and might be only weakly related to the physical characteristics of a product or service. For example, the Balloon Dog Orange, a 307.3 x 363.2 x 114.3 cm (121 x 143 x 45 inches) stainless steel statue by the American artist Jeff Koons, was sold at a 2013 Christie's auction for $58.4 million. The 2008 Urban Satchel Bag by Louis Vuitton, a handbag for women, made of fine Italian leather with urban motifs such as cigarette packages and advertised as "the world's most expensive handbag" ($150,000), reportedly quickly sold out the two dozen copies made.[10]

Let's make another consideration whose importance will become clear shortly. From the point of view of the theory of money, the most important function of markets is to determine prices. Markets establish prices as relative prices. A supercar might cost $500,000 while a cheap handbag might cost less than $10 at the street market. It is tempting to say that markets value a supercar at 500,000/10 = 50,000 times a cheap handbag. As observed above, Beinhocker counted tens of thousands of billions of comparisons possible in a modern economy. In practice, however, comparisons occur in restricted universes. Someone who buys a supercar will not likely buy a $10 handbag at a street market but a Hermès Kelly bag priced at upwards of $5,000.

How do comparisons extend to the entire market? We suggest that they do so by marginal interactions between sub-economies. Here's why. The people who buy $500,000 supercars and Hermès Kelly bags and travel in private jets live in a subset of the economy. However, these very wealthy individuals employ highly paid professionals and use expensive services so that there are interactions at the periphery with other sub-economies which might still be wealthy though less wealthy. This process extends to all subsets that together comprise the entire economy. In other words, we might think of a modern economy as formed by a number of sub-economies that have only limited interaction among themselves. Separation between sub-economies is a question of the level of income and wealth as well as other differences such as cultural, religious, or ethnic. One might object that economies are essentially a continuum of income, culture, and taste. But this is not empirically the case. People are divided into clusters. Clustering in similar groups extends to lifestyles. The type of lifestyle that one can afford tends to fall into a relatively small number of categories characterized by specific types of consumption such as private planes. Jumping from one category to another requires big increments in income or major cultural change.

It should be clear at this point that the notion of a single price level for an entire economy is not tenable. Relative price comparisons work within groups with similar income and wealth and similar profiles but do not work across groups with substantially different income levels and profiles. The point is that prices are only

partially related to characteristics that can be measured; they are also determined by immaterial factors. These considerations will become critical as we turn our attention to the discussion of inflation and are ultimately forced to conclude that there is no true measure of inflation and consider the consequences for our science of economics.

Notes

1 The asse or *aes* was a bronze coin used during the Roman Republic, the *modius* was a dry measure equivalent to 1.98 gallons.

2 See http://press.vatican.va/content/salastampa/en/bollettino/pubblico/2016/03/05/160305b.html.

3 The New York Comic Con is an annual New York City convention dedicated to comics, graphic novels, anime, manga, video games, toys, movies, and television. Reference here is made to the event of October 11, 2015.

4 A Ponzi scheme is a fraudulent investment scheme where high returns are promised to original investors at little risk with these returns being financed, not by profitable legitimate investments as marketed, but by a constant flow of new investments. Like with pyramid schemes, when the flows run out, the scheme collapses.

5 For the full-year 2016, Standard & Poor's (2017) announced that firms in the S&P 500 spent $536.4 billion on buybacks, while for the seven-year period 2009–2016 S&P 500 firms repurchased $2.75 trillion in stock. Another figure comes from Birinyi Associates, who estimate that U.S. listed firms had spent about $6.1 trillion buying back their own shares during the 11-year period 2005–2016.

6 Fichtner, Heemskert, and Garcia-Bernardo (2017) mapped the ownership of the "Big Three" passive index funds in the U.S. – BlackRock, State Street, and Vanguard – and found that they constitute the largest shareholder in 40 percent of all listed U.S. corporations and 88 percent of the firms in the S&P 500 index.

7 See *The Telegraph*, November 5, 2008.

8 See *The Telegraph*, June 9, 2017.

9 One might object that a stock is ultimately a title of ownership of a company in its complexity. But this is not true. A company is a legal person; the owner of a stock owns only the fraction of the cash generated by the company corresponding to the value of the stock. Stockowners might have the right to participate in shareholder meetings and request changes of strategy or management but they do not own the assets of a firm. Firms are legal persons; they own assets, stockholders own only shares issued by firms.

10 We might be inclined to think that the idea of putting a high price tag on an item of conspicuous consumption to make it more desirable is an invention of modern marketing and its studies on human behavior. Actually, it dates back to the 4th century BC: Aristotle recognized, in Book 1 of his *Politics*, that the "use value of a good or service will be increased if it can be consumed conspicuously." From Younkins (2005).

References

Aly, Samuel. 2017. "The Gracchi and the Era of Grain Reform in Ancient Rome." *Tenor of Our Times*: vol. 6, Article 6. Available at: http://scholarworks.harding.edu/tenor/vol6/iss1/6.

Azevedo, F.A., L.R. Carvalho, L.T. Grinberg, J.M. Farfel, R.E. Ferretti, R.E. Leite, F.W. Jacob, R. Lent, and S. Herculano-Houzel. 2009. "Equal Numbers of Neuronal and Nonneuronal Cells Make the Human Brain an Isometrically Scaled-up Primate Brain." *Journal of Computational Neurology*, vol. 513, no. 5 (April 10):532–541.

Bak, Per, Simon F. Nørrelykke, and Martin Shubik. 1999. "Dynamics of Money." *Physical Review E*, vol. 60, no. 3:2528–2532.

Barro, Robert J. and Xavier Sala-i-Martin. 2003. *Economic Growth*. New York: McGraw-Hill.

Beinhocker, Eric D. 2007. *The Origin of Wealth; The Radical Remaking of Economics and What It Means for Business and Society*. Boston, MA: Harvard Business Review Press.

Benedict, Ruth, 1946. *The Chrysanthemum and the Sword: Patterns of Japanese Culture*. Boston, MA: Houghton Mifflin.

Cauwels, Peter and Didier Sornette. 2012. "The Illusion of the Perpetual Money Machine." The Swiss Financial Institute Research Paper, no. 12–40. (October 27).

Chandler, Alfred D. 1977. *The Visible Hand*. Cambridge, MA: Harvard University Press.

Childe, Gordon. 1982 [1942]. *What Happened in History*. London: Peregrine Books [Penguin Books].

Coase, Roland. 1937. "The Nature of the Firm." *Economica*, New Series, vol. 4, no. 16 (November):386–405.

Daly, Patricia A. 1981. "Agricultural Employment: Has the Decline Ended?" *Monthly Labor Review*, vol. 104, no. 11 (November): 11–17.

Fama, Eugene. 1970. "Efficient Capital Markets: A Review of Theory and Empirical Work." *The Journal of Finance*, vol. 25, no. 2:383–417.

Fichtner, Jan, Eelke M. Heemskert, and Javier Garcia-Bernardo. 2017. "Hidden Power of the Big Three? Passive Index Funds, Re-Concentration of Corporate Ownership, and New Financial Risk." *Business and Politics*, vol. 19, no. 2 (June):298–326.

Grantham, Jeremy. 2017. "This Time Seems Very, Very Different." GMO Quarterly

Letter (1Q).
Kurzweil, Ray. 2013. *How to Create a Mind: The Secret of Human Thought Revealed*. Harmondsworth: Penguin Books.
Langlois, Richard N. 2002. "The Vanishing Hand: The Changing Dynamics of Industrial Capitalism." *Economics Working Papers*. 200221. http://digitalcommons.uconn.edu/econ_wpapers/200221.
Malinowsky, Bronislaw. 1922. *Argonauts of the Western Pacific*. New York: Routledge.
Mauss, Marcel. 1923. «Essai sur le don: Forme et raison de l'échange dans les sociétés archaïques.» Année Sociologique [reprinted in *Sociologie et Anthropologie*. 1968. Paris: PUF, Collection Quadrige, 4e éd.].
Minsky, Hyman P. 1986. *Stabilizing an Unstable Econ omy*. New Haven, CT: Yale University Press.
Pilkington, Philip. 2014. "Endogenous Money and the Natural Rate of Interest: The Reemergence of Liquidity Preference and Animal Spirits in the Post-Keynesian Theory of Capital Markets." Levy Economics Institute. Working Paper No. 817.
Robertson, Denis Holme. 1923. *Control of Industry*. New York: Harcourt, Brace and Company.
Shubik, Martin. 2010. *The Theory of Money and Financial Institutions*. Cambridge, MA: MIT Press.
Standard & Poor's. 2017. "S&P 500 Buybacks Total $135.3 Billion for Q4 2016, Decline for Full-Year 2016." *PRNewswire*. (March 22): www.prnewswire.com/news-releases/sp-500-buybacks-total-1353-billion-for-q4-2016-decline-for-full-year-2016-300427467.html.
Turner, Adair. 2013. "Credit, Money and Leverage: What Wicksell, Hayek and Fisher Knew and Modern Macroeconomics Forgot." Speech made at conference *Towards a Sustainable Financial System*. Stockholm School of Economics. (September 12): https://cdn.evbuc.com/eventlogos/67785745/turner.pdf.
Werner, Richard A. 1992. "Towards a Quantity Theory of Disaggregated Credit and International Capital Flows." Paper presented at the Royal Economic Society Annual Conference, York, April 1993, and at the 5th Annual PACAP Conference on Pacific-Asian Capital Markets, Kuala Lumpur, June 1993.
Werner, Richard. 1997. "Towards a New Monetary Paradigm: A Quantity Theorem of Disaggregated Credit, with Evidence from Japan." *Kredit und Kapital*, vol. 30, no. 2 (July):276–309.
Younkins, Edward W. 2005. "Aristotle and Economics." *Le Québéquois Libre*. (September 15): www.quebecoislibre.org/05/050915-11.htm.

2

The Theory of Money

Basic concepts part II

In Chapter 1, we outlined the social and economic framework for the theory of money. Summarizing our discussion, money is a social tool based on the notion of ownership and exchange. Complex economies and societies without money can exist, but their functioning would require significant changes with respect to our market economies, in particular a reduced role for ownership and a larger role for rules. Our discussion, of course, is centered on market economies with money. No developed market economy is truly a market economy but a mix of free markets and planned sub-economies. Let's start our discussion here with looking at if an economy society can work without money.

2.1 Do we need a theory of money?

Money plays a fundamental role as the vehicle for transmitting and implementing economic decisions. But does this call for a theory of money? The answer is "Yes." Money is not simply a mirror of the real economy; it is "manufactured" in determined amounts, selectively distributed, and allows the transmission of economic decisions. Money has a dynamics of its own that shapes the power structure in any money-based society. This is why we need a theory of money. More precisely, we need to integrate a theory of money and economic theory.

It is striking to read how kingdoms and empires were built on the availability of precious metals such as gold or silver to pay for the expenses of the sovereign's lifestyle and war. An important source of financing for Alexander the Great's wars was the Macedonian gold and silver mines from which coins were minted though, according to Plutarch, the most important source of financing was short-term loans of 1,460 talents (a talent was approximately equivalent to 26 kg of silver).[1] It might be a total coincidence but it is interesting to note that the title of the Holy Roman Emperor passed from the Carolingians to the Ottonians[2] in 962 when Pope John XII crowned Otto the Great emperor. At around the same time silver mines were discovered in Saxony, near Goslar. The mines funded the kings – and above all their large army – for two centuries. The industrial and military power of Great Britain in the 19th century has often been attributed to its efficient financial system.

Armies are the purest examples of command societies[3] but they need a market economy to satisfy logistic and manufacturing needs. As argued by Coase (1937) and his followers and discussed earlier, inside a society, there is almost invariably a mix of command and free-market segments in the economy (see Section 1.2.2). They suggest that a mixed free-market/command economy is more efficient than either a purely free-market economy or a purely command economy.

We adhere to the idea that money is not neutral and that changes in the money supply have economic effects, though measuring these effects is challenging. As we will observe in our discussion on inflation indices in Chapter 6, there is no unique definition of "real" economic quantities as opposed to "nominal" quantities. The effects of the quantity of money on the economy might be very different in function of how the money is generated and distributed. For example large amounts of the money generated might not reach the bulk of the economy (Main Street) but remains concentrated in sectors such as financial markets (Wall Street). If we measure inflation with an index biased towards the consumption of a person with an average income, we might find that the newly generated money has no effect on real economic

variables, that it has not produced inflation, but has simply followed the natural growth rate of the economy. However, if we take into account the true complexity of the economy and follow the channels through which the money has flowed, we might reach a different conclusion.

Going forward, governments and central banks will have to look at the theory of money as the theoretical underpinning of how to allocate means of payment to individuals and entities. The discussion of possible new means of payment and new ways of allocating the means of payment goes well beyond the purely technical side. Access to means of payment is access to economic power – and the right to survive. The technical side of money and the political economics that decides what type of society we will have go hand in hand.

2.2 Operationalism and theories of money

Modern science is rooted in the principles of operationalism. Operationalism requires that the terms used in a theory be firmly rooted in observations. So any theory of money that aspires to be considered a (near-)scientific theory must adhere to the principles of operationalism. Mainstream economic theory is, however, not based on operational principles in that it is not based on observations, but makes use of presumed self-evident a priori concepts. Examples include the concepts of inflation and utility.

Box 2.1 Operationalism

Operationalism is a methodological principle proposed by Percy Williams Bridgman, recipient of the 1946 Nobel Prize in Physics, in his book *The Logic of Modern Physics* (1927). The key tenet of operationalism is that all variables of physics are defined by the operations for observing and measuring them.

There is no place in physics for apparently intuitive a priori concepts. A classic example is the concept of simultaneous events. The concept of simultaneity seems to be a well-defined a priori concept. But the Theory of Relativity has shown that simultaneity is a concept relative to the system of reference of the observer. Two events can be simultaneous for one observer but not simultaneous for another observer. The relativity of the concept of simultaneous events is a consequence of the operations that we perform to observe simultaneity and the invariance of the speed of light for observers in mutual movement.

Operationalism is a methodology of empirical science rooted in the process of theory formation typical of physics. Operationalism is also related to the philosophical movement called Logical Positivism. Logical positivism was proposed in the first half of the 20th century by a group of philosophers – the Vienna Circle – who used to meet at the Café Central in Vienna. Prominent members of the group were Moritz Schlick, Otto Neurath, and, above all, Rudolf Carnap. Richard von Mises, the younger brother of the economist Ludwig von Mises, was also a member of the Vienna Circle. Ideas similar to those of the logical positivism of the Vienna Circle were proposed at about the same time in Berlin by Hans Reichenbach and in London by Alfred Jules Ayer.

Logical positivism is a rather radical philosophical theory. It claims that the

meaning of any assertion is in its verification. But ultimately verification is based on sensorial experience so that the meaning of any sentence is rooted only on perceptions. Rudolf Carnap's (1928) *The Logical Structure of the World* is a monumental work that tries to show how we can construct all knowledge starting from experience with the help of logic only.

Logical positivism has been enormously influential. It was later criticized by philosophers such as the German Carl Hempel on the basis that scientific knowledge requires abstract terms such as force, mass, and field. Leaving aside the philosophical debate as to whether knowledge can be logically constructed from experience, the principles of operationalism remain valid in the physical sciences. It is true that science uses highly abstract terms in its formulation but ultimately scientific theories are, globally, based on experience.

The Copenhagen interpretation of quantum mechanics, proposed by Niels Bohr in 1931, explicitly states that the equations of physics – and quantum mechanics in particular – are simply models that connect observations. The nature of observations remains the subject of serious debate even in physics.

Let's introduce a subtle but important point. A clear distinction should be made between variables that are implicitly defined through a scientific theory and the "hidden variables" of economics such as volatility. A variable defined through theory is a variable whose observation depends on the theory. Consider, for example, a physical concept such as temperature. While we cannot directly observe temperature, we can observe phenomena such as the elongation of a column of mercury or a reading from a digital thermometer based on the variation of electric resistance of some special material in function of the temperature. Temperature is a theoretical term of thermodynamics and is related by the thermodynamic theory to all other physical phenomena. Temperature is thus an important variable because it is linked to all other physical variables through the theory itself.

Now consider economics and finance theory. A hidden variable is a variable that is not observed directly but which can be computed from other variables. Consider inflation. Current macroeconomic models are intended to model real quantities such as real prices and real output as opposed to nominal quantities. But we cannot observe these real quantities. Actually real quantities such as real GDP are obtained by discounting nominal quantities by an inflation index, but, as we will see in Chapter 4, no index can claim to be the true index. If an index *ITrue* could be placed in a global theory then it would be possible to argue that *ITrue* is at least a useful index. Consider the well-known Quantity Theory of Money in the form: $PQ = MV$. In

this equation MV can be observed, at least in principle, while P is the non-observable price level. If we could find some consumer price index based on observed prices of consumer goods and find that it coincides, up to acceptable noise, with the P of $PQ = MV$, then we would have both a macroeconomic theory and an operational definition of inflation.

Note that it would be futile to ask if I is the *true* definition of inflation. The variable I would be a theoretically useful variable that we could call inflation. We cannot ask for more. Of course we can still claim that indices such as the Laspeyres or Paasche indices (see Chapter 6 for a discussion on these and other indices) are useful, but we have to specify in what sense they are useful.

It might be objected that we can still develop a lot of theory without specifying any inflation index. Mathematically this is true, but we are concerned with empirical theories, applicable to the real world, not with idealized mathematical constructions.

As for the utility function, as it stands today, utility functions are simply mathematical abstractions used to represent the decision-making process of economic agents. They are typically aggregated to form the utility function of the representative agent. But utility functions are not based on observations and cannot be empirically estimated. An economic theory that takes basic concepts such as inflation and utility functions as a priori concepts cannot be considered an empirical theory. A theory then must explain a universe of phenomena and predict the future behavior of the object of the theory, in our case money.

2.3 The concept of stock-flow consistency

Stock-flow consistency is a concept known in corporate accounting as double bookkeeping. By "stock" is meant the amount of some quantity such as money that is owned. In the accounting of any firm, financial flows must match changes in assets and liabilities. Similar principles apply to the modelling of any economy. Historically, the idea of stock-flow consistency in economic modelling goes back to Morris Copeland's (1952) "flows of funds." Copeland studied the flows of money in and out of different sectors and laid the basis of stock-flow consistent models. Skeptical of the current state of economics, Copeland believed that the discipline must take into account the institutions prevailing in each moment. He advocated the study of economics as an empirical science, collecting and analyzing data.

Copeland (1949) was explicit about the sort of problems that he thought economists should address: When the total purchases of our national product increase, where does the money come from to finance them? When purchases of our national product decline, what becomes of the money that is not spent? What part do cash balances, other liquid holdings, and debts play in the cyclical expansion of money flow?

Copeland's work was brilliant and innovative but failed to change the direction of macroeconomics. Two contemporary economists – Winne Godley and Marc Lavoie (2007) – took up Copeland's ideas and proposed modelling stock-flow consistency where changes in stock variables are always equal to the sum of the relative flows. Stock-flow accounting is important when we want to introduce a financial system in economic modelling. At this point it becomes critical to understand how the flows of money match changes in economic variables and how they can eventually lead to a financial crisis. Macroeconomic models with a financial system and money that respect stock-flow consistency are discussed in Chapter 8.

2.4 Money and macroeconomics

The effect of money on the real economy and how the theory of money must be incorporated into macroeconomics are the subject of Chapter 8; here we offer a preview of the main arguments. As already observed, the ability of banks to create almost unlimited quantities of money has profound effects on the real economy. Money transmits economic power by conferring the ability to purchase goods and services. Following the Quantity Theory of Credit first articulated by Richard Werner (1992, 1997), we have to divide the newly generated money into two parts: one used for GDP transactions and the other used for financial transactions. The former finances production and eventually consumption; the latter finances asset purchases.

According to Minsky's analysis (1986), money used for financial transactions might produce asset inflation and eventually lead to financial crises; money used for GDP transactions flows unequally through the economy and serves to segment the economy into subsectors with different levels of income and wealth.

Macroeconomics and the theory of growth are profoundly influenced by money. In classical theories of growth, growth is real growth in terms of the quantities, number, and quality of products. But in real economies this physical growth cannot be directly assessed as there is no way to physically aggregate products and services. We therefore aggregate through prices, creating nominal GDP and nominal growth. Nominal quantities are then discounted by inflation rates to yield real quantities. But as seen we can establish neither a meaningful price level, nor, consequently, an inflation rate for the entire economy.

The financing of the state is another critical issue where the theory of money meets macroeconomics. The current way of financing the activities of modern states is through taxation and the issuance of debt. The issuance of state debt is a controversial issue that led to the much-debated austerity programs in the period following the 2008 financial crisis and ensuing Great Recession. In particular, the proponents of Modern Money Theory (MMT) favor an approach more in line with Abba Lerner's (1943) functional finance. Lerner thought that the activity of the state should be determined by the availability of real resources and not by financial considerations. (This issue will be discussed in more detail in Chapter 8.)

Box 2.2 The Naïve Theory of Money and the Fallacy of Composition

For non-technical readers, the theory of money as bank deposits that can be created and destroyed requires some explanation. The key point is what economists call the Fallacy of Composition: What is true for an individual is not true for the economy as a whole.

Many today still hold onto the notion that money is something real and, once created, permanent. This was the case when money was made of gold or silver coins but it is no longer the case with the modern banking system.

Most people today have bank accounts in which their "wages" and other gains are stored and from which they withdraw banknotes, and against which they write checks, or make electronic payments of one sort or another. Money flows in and out of bank accounts, the objective being to ensure that the flows in cover the flows out – unless of course one takes out a loan for one reason or another, in which case the flows in must cover the repayment schedule. The existence of money is taken for granted.

But to arrive at an understanding of the key issues associated with money calls for some effort of the imagination. The first effort is to aggregate all payments received by individuals, be they wages or remuneration for professional services, rents of one kind or another, dividends, or other. Consider a given day and all persons who received some sort of payment (inflow). Although most people find it natural that they receive these inflows for one reason or another, in aggregate, one might ask: Where does all this money come from? Well, most people would agree that money that flows into their bank accounts comes from firms that employ them or pay them dividends and, in the case of professional services, from persons who have engaged their competences (unless it is a question of inheritance or other).

But how is it that firms, for example, have money to pay for labor or goods? Some reflection would lead us to conclude that firms have money because they have sold their products or services. But to whom? Excluding for simplicity import/export, in aggregate firms sell their products and services back to employees and others who have a professional activity or who live off rents and profits.

At this point the reader has understood that in aggregate money flows in circuits, from firms to employees as wages and dividends and back to firms through the sale of products and services. What we have is a big circular flow of money.

Now let's make another effort of the imagination. Consider again a closed economy such as a modern nation without international relationships. (International exchanges would complicate the exposition without adding

anything to our understanding of money.) Consider all the individuals and firms in this nation. Most individuals and firms have one or more bank accounts where they keep the bulk of their money. They might also keep some coins and banknotes in their wallets and homes.

Consider now the set of all bank accounts and the set of all wallets, bank vaults, ATMs, and all other places where banknotes can be stored. What are financial flows? Essentially movements from one bank account to another or physical transfer of banknotes from one place to another. Note that a bank account is ultimately a computer record of the amount of money in the account. Banknotes can be withdrawn or deposited, a bank transfer made, a check written or deposited, increasing/decreasing one account and decreasing/increasing another account.

So, ultimately, financial movements today consist in changing the computer record of the amount in bank accounts or physically transferring banknotes. For most people this snapshot would offer a sufficient theory of money. When we receive our wages or other gains, we do not question the nature of money. We simply verify that after receiving our due inflows our account has increased by a corresponding amount; we further trust that we can write checks, order bank transfers, withdraw banknotes, and so on. From the point of view of the user, the entire banking system can be represented as a set of bank deposits (computer records) that change in time plus the physical movement of banknotes.

But let's make another effort of the imagination. We can ask how all this money was initially created. You might say that the current situation is the result of a history of many different activities. But this explanation is inconclusive. If we have a good, well-paid job our bank account will increase in time unless we overspend. However, in aggregate, the sum of all bank accounts and banknotes in circulation does not change. In fact, any payment received by one party is disbursement for another party. The fact that payments imply that one account is debited, that is reduced, and another credited, that is increased, by an equal amount holds the total constant.

Under the current assumptions, if we go back in time, we find the same situation as today. We find a large number of accounts with deposits whose value is individually different from current accounts but whose sum (in aggregate) is the same. To make some progress we have to expand the analysis beyond payments. Suppose that client Z wants to buy a house but does not have enough savings and has no other property that he or she can sell. Most likely he or she goes to the bank and asks for a loan. If Z is considered creditworthy, the

bank will make a loan that engages Z to pay a periodic amount of money to the bank. After granting the loan, the bank will credit Z's account with the amount of the loan, that is, it will increase the value of Z's account by the amount of the loan.

But where does the money of the loan come from? From nowhere, because in making a loan to Z, the bank has created money *ex nihilo*. The bank does not lend its capital. To be persuaded, look at a bank's balance sheet. The amount of outstanding loans is many times the amount of capital of any bank.

Nor do banks (any longer) lend the money of their clients; they simply create a deposit when they grant a loan. So the loan extended to client Z is created *ex nihilo* and the amount adds to the total of deposits. Client Z will likely use the loan to pay for the house he or she intends to buy. The money will therefore leave Z's bank and enter into circulation.

By granting a loan to a client, money has been created and the sum of all deposits has increased.

When the first loan repayment is due, the bank will receive the money from Z and reduce the outstanding debt of Z by a corresponding amount. What happens is that the bank account of Z is reduced by an amount equal to the first repayment. Actually this repayment can be divided into two parts: the repayment of the loan and the payment of interest. With the repayment of the loan, money disappears; we say that it is destroyed. Each time Z makes a payment, an equal portion of the money created is destroyed. The payment of interest, on the contrary, is transferred to the bank as revenue.

2.5 A framework for understanding theories of money

Though the many theories of money that have been proposed all have money as their key topic, the different theories actually deal with different phenomena. For example, the generation of money as credit has nothing to do with the generation of fiat money. In addition, the notion and the mechanics of money are in a constant state of evolution. To understand the different views on money and their possible evolution, we adopt a common framework of analysis based on five points: (1) the nature of money, (2) how money is generated, (3) how money acquires value and how its value changes over time, (4) how money is distributed, and (5) how money interacts with the real economy. All theories of money deal with one or more of these key aspects which we will briefly present here and develop in the following chapters. We will try to separate as much as possible the purely theoretical, cognitive, aspects of the theories of money from the debates regarding the consideration of different economic and monetary policies.

2.5.1 The nature of money

The nature of money, and in particular the question as to whether money is something with an intrinsic value or something that acquires value by social convention or by government edict, have been discussed since Antiquity. For Alexander the Great's tutor, Aristotle:

> money is a medium of exchange that makes exchange easier by translating subjective qualitative phenomena into objective quantitative phenomena. Although subjective psychological want satisfaction cannot be directly measured, the approximate extent of want satisfaction can be articulated indirectly through money. Not only does money eliminate the need for a double coincidence of wants, it also supplies a convenient and acceptable expression for the exchange ratio between various goods. Money, as an intermediate measure of all things, is able to express reciprocity in accordance with a proportion and not on the basis of a precise equal ratio. Money, according to Aristotle, has become a convention or type of representation, by which all goods can be measured by some one thing. Money, as a modulating element and representation of demand, becomes a useful common terminological tool in the legal stage of the bargaining process.[4]

Discussions on the nature of money continue unabated today as, to the list of established forms such as coins, banknotes, and bank deposits, we add digital money and cryptocurrencies such as Bitcoin. In addition, we can add to money liquid assets such as money market funds, frequently referred to as near-money. But despite all the forms of money and the sometimes subtle theoretical issues in understanding the different natures of money, money is in general understood as purchasing power in the sense that it enables its holder to buy goods and services. The critical issue is

how money is created and distributed.

2.5.2 The creation of money

Coins are created through a physical process of minting; banknotes are printed. No problem here. The debate is on the generation of money as credit and, more recently, the generation of digital currencies. It is important here to distinguish the process of generation from the process of allocation of credit money (banknotes pass through the banking system). The classical theory of the multiplier states that the banking system can only multiply exogenous deposits. This view is currently challenged by the endogenous theory of money which argues that the banking system does not need deposits as it can simultaneously create deposits and loans and ask for reserves from the central bank. The competing exogenous (verticalist) and endogenous (horizontalist) theories of money have important implications for central banks' objectives in managing money, the central banks' objectives themselves being subject to debate. If we accept that money is endogenously generated by banks, what is the role of central banks now and in the future?

Both the verticalist and the horizontalist approaches consider simultaneously the generation and allocation of money. Strictly speaking, banks create new money by writing a number in a computer position that represents the deposit of a client. The question is: Why is this process accepted and credible? The answer, today, resides in how money is allocated. In aggregate, in current financial systems, new positions are created when clients take out loans or, eventually, when clients sell assets to the central banks via an intermediary commercial bank. Transfers between clients clearly do not create new money.

2.5.3 The distribution (or allocation) of money

The notion of "injecting money into the economy" assumes that someone gets the newly generated money, and with it the ability to transmit economic decisions. Allocating money implies allocating economic power. Currently, commercial banks are the principal allocators of new money through the granting of loans. This is, in itself, a very important characteristic of modern financial systems that might be subject to some revision in the future. For example, faced with automation making humans redundant, weak demand, and stagnant economies, the uniform distribution of money is being seriously discussed and experimented with.

2.5.4 How money is accepted, how it acquires value, and how its value changes over time

Lastly, there are the questions of how a specific form of money is accepted, how it acquires value, and how its value changes over time. There are two competing theories to explain how money gains acceptance. One suggests that money is accepted by social convention, the other – the state theory of money – argues that money is imposed by the state.

Today, one is more inclined to accept the state theory of money, first articulated by German economist Georg Knapp (1924). Knapp argued that state-issued money is accepted because the state requires it in payment of taxes and other sums due and, in turn, the state uses state-issued money to pay for the goods and services it requires. Note that in large developed countries, the government's share of the total GDP is typically somewhere between 40% and 50%. In 2015, those countries where the government's share of the total GDP was the lowest include the U.S. (37.7%) and Japan (39.5%) while those countries where the government's share was the highest were France (56.6%) and Finland (57%); Germany is in between with government spending representing 44% of the total GDP.[5] As almost everyone and every commercial activity is subject to taxation, individuals and entities must obtain state-issued money to pay taxes and other levies to the state. Citizens of the euro area, for example, must pay their taxes in Euros.

As for the theory that money is accepted by social convention, consider modern financial systems. Money as bank deposits is generated by commercial banks. But bank deposits are not state money though they might be denominated in state money. Credit money is accepted because the banking system has a credibility of its own. People trust the banking system for a number of reasons including that the system is backed by a strong central government and/or a central bank. Perhaps both the state theory of money and the notion of trust play a role in the acceptance of a specific money.

But there is another question related to the value of money intended as purchasing power. We know that in a market economy prices are formed through the intersection of supply and demand. But such intersections are implemented through money. Intuition informs us that prices are established in function of the quantity of money available. However, formulating this intuition in precise economic terms is a difficult problem only partially resolved. Due in particular to the inequalities in the allocation of money, and to the difficulty in defining a price level, there is no solid theory that can forecast changes in prices due to changes in the money supply.

Notes

1 From "Alexander the Great: A Very Competent Expert in Finances: Rate of Wages, Tax Reform and Financial Scandals." November 30, 2012:www.archaeology.wiki/blog/opinion/alexander-the-great-a-very-competent-expert-in-finances/.

2 The Ottonian dynasty was a Saxon dynasty of German kings (919–1024) named after its first Emperor Otto I. The Ottonian rulers were successors of the Carolingian dynasty in East Francia and displaced the Carolingians as heads of the Holy Roman Empire.

3 Following our definition of economies, armies are not economies because their objective is not to manufacture goods and services for consumption.

4 This summary of Aristotle's comments on the nature of money was made by Edward W. Younkins, professor of accountancy and business administration at Wheeling Jesuit University in West Virginia and author of *Capitalism and Commerce*. From Edward W. Younkins (2005).

5 See figures from the Organisation of Economic Co-operation and Development (OECD https://data.oecd.org/gga/general-government-spending.htm.

References

Bridgman, Percy. 1927. *The Logic of Modern Physics*. New York: Macmillan.

Carnap, Rudolf. 1928. *Der Logische Aufbau der Welt*. Leipzig: Felix Meiner Verlag. English translation by Rolf A. George. 1967. *The Logical Structure of the World. Pseudoproblems in Philosophy*. Berkeley: University of California Press.

Coase, Roland. 1937. "The Nature of the Firm." *Economica*, New Series, vol. 4, no. 16 (November):386–405.

Copeland, Morris. 1949. "Social Accounting for Moneyflows." *The Accounting Review,* vol. 24, no. 3 (July):254–264.

Copeland, Morris. 1952. "A Study of Money Flows in the United States." NBER Working Paper.

Godley, Winne and Marc Lavoie. 2007. *Monetary Economics: An Integrated Approach to Credit, Money, Income, Production and Wealth.* New York: Palgrave Macmillan.

Knapp, Georg Friedrich. 1924. *The State Theory of Money*. London: Macmillan and Co.

Lerner, Abba. 1943. "Functional Finance and the Federal Debt" in *Selected Economic Writings of Abba Lerner*. 1983. New York: New York University Press.

Minsky, Hyman P. 1986. *Stabilizing an Unstable Economy*. New Haven, CT: Yale University Press.

Younkins, Edward W. 2005. "Aristotle and Economics." *Le Québéquois Libre*. (September 15): www.quebecoislibre.org/05/050915–11.htm.

3

What is Money?

This chapter discusses the nature of money. After some historical remarks, we discuss the various forms of money: commodities, coins, banknotes, credit money, and, lastly, new forms of money including near-money and the shadow banking system, and alternative currencies including digital currencies and local currencies. The concept of credit money – a concept that lends itself easily to much confusion – is discussed, distinguishing credit money (1) as money redeemable into banknotes, (2) as purchasing power, and (3) as bank deposits associated with loans. The chapter closes with a discussion of two competing theories on the general nature of money, the Metallists (who argue that money must have intrinsic value) and the Chartalists (who argue that money is fiat money).

3.1 Some brief remarks on money throughout history

The historical perspective on money found in textbooks is that money evolved from barter. Paul Samuelson (1976, pp. 274–276), for example, describes the development of money in several stages, from barter to commodity money, to paper money, and lastly to credit money (bank money in the terminology of Samuelson). The mathematical models of coins discussed in Chapter 4 adopt the same basic perspective.

Textbooks typically describe primitive societies without money where goods are exchanged through barter. Note that the notions of ownership and exchange must have developed more or less in parallel, possibly in forms different from the current notion of ownership. But barter might be inconvenient: it requires that the bartering parties want to exchange things of the same value, the "double coincidence of wants," a term first used by Stanley Jevons (1875). As is explained in textbooks, in order to overcome the fundamental limitations of barter, societies began to adopt divisible objects such as grain, cowry shells, or some form of metals as a medium of exchange and/or standard of value. An early example of this comes from Bronze Age Sumeria where the urban revolution led to an increase in the exchange of goods and services which in turn created the need for a common standard for measuring and valuing goods. The first widely accepted standard there was barley, the basis of Sumerian life. In the Early Dynastic period (2900–2350 BC), though wages and rents were still mostly paid in barley, units of metal – silver and, for small sums, copper – were already the most generally accepted medium and standard of exchange; they were to remain so for the next 2,000 years in Mesopotamia. With the adoption of a conventional metallic standard, the transition from a natural economy based on barter to a money-based economy had been made.[1]

3.1.1 Barter

Before proceeding with the evolution of "money," let's step back a minute and see how the role of bartering in early civilizations has been questioned by anthropologists. David Graeber (2011), for example, notes that anthropologists have not found evidence of societies based on barter; he puts the accent on debt. Graeber writes:

> When economists speak of the origins of money, for example, debt is always something of an afterthought. First comes barter, then money; credit only develops later . . . For almost a century, anthropologists like me have been pointing out that there is something very wrong with this picture. The standard economic-history

> version has little to do with anything we observe when we examine how economic life is actually conducted, in real communities and marketplaces, almost anywhere – where one is much more likely to discover everyone in debt to everyone else in a dozen different ways, and that most transactions take place without the use of currency.
>
> *(p. 22)*

Graeber cited descriptions by the 19th-century American anthropologist Lewis Henry Morgan of the Iroquois who stored goods in common in a sort of warehouse from which the goods were successively allocated to members of the community by women's councils, without ever being traded (p. 29).

Graeber and others suggest that the most primitive forms of exchange were based, not on barter, but on debt and credit, the various ways of recording debts and credits being specific to a society. A widely cited example comes from Yap, an island in the western Pacific Ocean whose legal tender was a large stone. The macroeconomist (and fund manager) Felix Martin (2013) describes how the Yapese conducted exchange using credit accounts and clearing by carving notches on large round stones called *fei*. Another example comes from Ancient Sumeria, already mentioned above, where wedge-shaped marks (cuneiform) were imprinted on soft clay tablets that were then fired in kilns to record the credit and debt of persons working for the temple.

Leaving the argument on barter to the anthropologists, it is interesting to note how easily a community that was accustomed to the use of money for exchange turned to barter when money disappeared from circulation. In the Tuscan hill town of Piancastagnaio, some 75 kilometers south of Siena, money disappeared from circulation after World War II. The town and surrounding countryside counted some 2,000 inhabitants, half living in the hill town and working in the nearby mercury mines and half in the countryside making a living as small landowners. The hill town dwellers lacked food, the landowners carbide to light their homes and allow work in the fields by night. So those that worked in the mines creamed off a bit of carbide from their lamps as they descended into the galleries and smuggled it to the countryside where they exchanged the carbide for grain and other foodstuffs. There was no fixed rate of carbide/foodstuffs exchange: it depended on one's bargaining position at any given time.[2]

Let's now return to our linear story of the evolution of money.

3.1.2 Commodity money

Commodity money is the first type of money whose nature we explore. Strictly speaking, we call "commodity money" any commodity used as a medium of

exchange. Hollywood films on the Wild West left generations of cinemagoers with the image of California gold miners strutting into local stores and throwing their gold dust or gold nuggets on the counter to pay for goods.

It is easy to imagine the use of commodities as money but economists and anthropologists now question its use historically. As pointed out by Goodhart (1998), there are serious problems in accepting commodities as a means of payment. If the commodity is a metal, one needs to assess the quality of the commodity, the purity of the metal, and be able to weigh or otherwise measure the quantity. If metal comes in bullion and not in powder, the exact amount of metal needs to be cut.

Commodities, however, can be used as convenient *measures* of value. With the decipherment of Sumerian cuneiform in the mid-nineteenth century, we learned a lot about the Sumerian economy. We now know, for example, that around the year 3500 BC the temple administrators developed a single uniform system of accountancy with the silver shekel as the basic monetary unit. Graeber (2011) gives the comparative value established for a shekel based on barley: one shekel's weight in silver equals one gur (or bushel) of barley. The shekel was subdivided into 60 minas, corresponding to an equivalent portion of barley. The shekel was not, however, a product of commercial transactions but a means for court bureaucrats to keep track of resources and transfer them between departments (p. 39).

More recently, in the middle of the 12th century AD, linen cloth of a standard size was still being used as a medium of exchange among the West Slavs (the Rani) settled on the island of Rügen, located off the Pomeranian coast in the Baltic Sea and now Germany's largest island. The use of the linen cloth as a medium of exchange among the Rani dates back to as early as the 10th century when the Jewish merchant Ibrahim ibn Jacub first reported its existence.[3]

3.1.3. Coins

The idea of coinage follows almost naturally from the use of metal as a commodity for exchange. As bullions evolved into minted coins, exchange became the exchange of coins for goods and services. Graeber (2011) associates the first minting of coins around the 6th century BC with what the German existentialist philosopher Karl Jaspers called "the Axial Age." Jaspers' Axial Age runs from about 800 BC with the Persian prophet Zoroaster to the 6th century BC when, more or less contemporaneously, the Greek culture produced the mathematician and mystic Pythagoras (570–495 BC), the Indian culture, the Buddha (563–483 BC), and China Confucius (551–479 BC) to close with what Jaspers called the "Spiritual Age"

centered on figures like Jesus and Mohammed (pp. 223–224). Graeber then connects the Axial Age to the first coinage, noting that the three parts of the world where coins were first used correspond to the very parts of the world where religious and philosophical creativity thrived, that is, the kingdoms and city-states around the shores of the Aegean Sea, in the Ganges valley in northern India, and around the Yellow River in China (p. 224).

Anthropological studies date the first minted coins to around 600 BC, in the Iron Age kingdom of Lydia on the Ionian Sea.[4] Conquering Lydian armies took as booty women and metals, including silver, gold, and electrum. Lumps of electrum, a naturally occurring gold–silver alloy, were melted down and used to pay soldiers' wages. Successively the lumps were hammered and stamped with an insignia by jewelers before coinage was transferred to a newly created royal mint. Coinage was successively adopted by Greek cities on the Ionian coast, by Greece itself, by Persia after it took Lydia in 547 BC, and by the Romans in the 4th century BC in imitation of Greek culture.

Many anthropologists believe that slavery and (mercenary) soldiers were behind the first large-scale coinage: slaves mined gold and silver; (mercenary) soldiers got paid with coins, in what Geoffrey Ingham (2004) called the "military-coinage complex." Struggles involving the Phoenician trading cities were to end in the mass enslavement of their populations by the victors, first the Persians in Sidon, then the Greeks of Alexander in Tyre, and lastly the Romans in Carthage. Of the tens of thousands of captives, many were sent to work in the gold and silver mines. The objective: to mine the metals needed for coinage to pay and provision troops. It was estimated by Glynn Davies (1996) that Alexander needed half a ton of silver a day just for the wages of his 120,000 troops and camp followers (p. 229). When the republican government in Rome was vacillating, the then-general Julius Caesar minted his own coins to pay his soldiers from the silver acquired through conquest.[5]

More in general, the Romans had a well-developed coinage system to support the needs of the economy. Though agriculture was at the base of the Roman world, the economy was quite complex.[6] Coins served to purchase goods and services from the various economic players, from prostitutes, bakers, and soldiers to small-scale manufacturers of building materials, tableware, tools, and textiles to merchants conducting long-distance trade with northern Europe, India, and China. Coins of brass, bronze, copper, silver, and gold were minted and circulated under strict rules for weights, sizes, value, and metal composition.

In post-Roman Frankish kingdoms, Guy Fourquin (1990) relates that money lost its unity at the end of the 6th century as private metal workers (officines privées) holding the metal were allowed to mint coins. It has been estimated that soon more

than 800 mints coined up to 5,000 different moneys. The first emperor of the post-Roman West, Charlemagne, reserved for himself the right to mint coins but had problems enforcing this: the Carolingian lands were too vast for his administration to exercise effective imperial control. By the year 900, royal coins gave way to feudal coins. Today the use of coins is relegated to small transactions such as those that occur at the local bakery and even there tends to disappear with the swapping methods of credit cards.

Considerations related to coins and coinage

If the value of coins is exactly the value of the metal they contain, then there is no direct profit in minting coins. Coinage effectively belongs to those who have metal; the question of the distribution of money is pushed back to the question of the distribution of the metal. If a tax on coinage (seigniorage[7]) is applied, then the process of minting becomes profitable for the government or the sovereign. Seigniorage can be applied even if the minted coins have a value identical to the value of the metal.

The value of the coin might not be directly linked to the value of the metal of which the coin is made. This situation is inevitable if there are two or more metals involved in the minting of coins and if there is a fixed ratio between the value of the different coins. However, even in monometallic systems, coins might have values different from their content in metal. Governments or sovereigns typically impose – or try to impose – the weight, fineness (metal content), insignia, and nominal value of each coin to be minted.

There are a number of issues associated with the use of coins. First, at the moment of minting, the purity of the metal must be assured. Today it is a simple matter to test the purity of gold but in Antiquity methods for testing gold were neither accurate nor reliable. The touchstone method, then widely used to test the purity of metals, works by scratching a piece of the metal to be tested on a special stone. The purity of the metal is gauged by the color of the imprint on the stone. Later methods included testing gold and silver against corrosive acids. Today the purity of gold is assessed with spectrographic methods.

As for the institutions involved in the minting of coins, the question is primarily one of the minting operations and of the legal tender. A mint might be a state-owned enterprise or a private enterprise licensed by the issuing authority. The right to mint might be reserved to the sovereign (or state) or shared with other powers such as high-placed members of a cult or nobles.

We say that an area provides free minting if anyone therein who owns bullion can have the bullion transformed into coins. But even in the case of free minting there was generally a charge. In most cases, in order to have bullion minted a person had to pay for the physical cost of minting plus a percentage called seigniorage to the governing body. The mint price is the price at which the mint is ready to buy gold or silver to convert it into coins. If the mint does not charge any cost for minting, then minting is said to be gratuitous.

Coins become legal tender if the issuing authority decrees that they must be accepted in payment for any transaction and in repayment of any debt. If there is no limit to the amount that can be settled with the coins, coins are said to be unlimited legal tender; if there are limits to the amount that can be settled with coins, coins are said to be limited legal tender. Small coins, for example, might be limited legal tender. A monometallic system is one in which only one metal is legal tender and is freely minted; a bimetallic system is one where both gold and silver are unlimited legal tender and there are no charges associated with minting; a *de jure* (by law) bimetallic system is one where both gold and silver coins are unlimited legal tender and the ratio is fixed by the law.

The Romans used various metals in minting coins, including gold, silver, bronze, and copper, but the most important coin was the silver denarius. In the late Roman Empire, gold coins progressively disappeared in the West so that sometime between the 5th and 7th centuries, they were no longer used in exchange but were limited to serving as a unit of account or store of value.[8] Thereafter, in medieval Europe, silver coins were the norm until the mid-13th century when Florence minted a gold coin – the Florin – which gained wide acceptance in long-distance trade across Europe. Silver coins, however, continued to circulate. In the 16th and 17th centuries and later up to the 19th century different systems were in place in different countries.

At first unmarked, coins were to become more sophisticated and benefit from the guarantee of an issuing authority. An illustration of this process comes from ancient Rome where the *aes rude*, a rectangular ingot of cast bronze used in the 5th and 4th centuries BC, evolved into the *aes signatum* which carried inscriptions specifying its weight and then in 289 BC became the *aes grave* whose value was specified by the Roman state.

However, metal coins were subject to debasement. The term *debasement* indicates the process by which the value of the metal that forms a coin becomes less than the nominal value of the coin. Debasement might be the design of the issuing government or the result of petty criminal action. A well-known example of debasement by the issuing government comes from Ancient Rome where the nearly pure silver denarius (successor of the *aes grave*) weighing about 4.5 grams was, by the second half of the

3rd century AD, reduced in size and purity by consecutive debasements until the coin contained only about 2% silver.

Problems of debasement and the stability of the value of coins did not, of course, end with the fall of Rome. Fourquin (1990) relates that in most of southern France and Italy the value of silver coins was so unstable in the second half of the 11th century until around 1140 that, in written acts, the equivalent of the coin was given in livestock and foodstuffs (pp. 139–140). In 16th-century England, Henry VIII ordered the debasement of coins to finance his lifestyle and wars. Known as the Great Debasement (1544–1551), the Crown's policy saw gold standards drop to 20 carat from 23 carat previously and the amount of silver reduced to 25% from 92.5%.

Petty criminal debasement involves removing small unnoticeable portions of the coin by shaving, clipping, or filing. Clipping was, along with counterfeiting, the main form of coin crime in 17th- and 18th-century England. Suspicious shopkeepers kept weighing scales close at hand.

3.1.4 The emergence of a banking system and credit

In Rome, as well as in many areas of the Mediterranean, early banking started in temples, considered safe places for the storage of value because they were managed by members of the cult and heavily guarded. As social structures in the Ancient World grew more complex and spanned ever wider territories, more sophisticated financial facilities were required to support trade and finance vast infrastructure works. New figures in banking emerged. The *argentarii* (from argentum, Latin for silver), for whom we have mention dating from 250 BC, were private individuals who offered banking services such as money changing, safe deposits, lending, and payment of commercial transactions – and also engaged in speculative finance. These early bankers belonged to a guild and performed their activity in stalls around the Forum, in shops, and in state-owned banks.

Other new figures in banking in Rome were the *mensarii* and the *nummularii*. As was typical of the times, law in ancient Rome authorized enslavement should a person fail to repay debts. To mitigate the consequences of insolvency in periods of economic stress, the Roman Republic (352 BC) appointed public bankers called *mensarii*, whose role was to offer state support to persons who could not repay debt but who could offer some guarantee. As for the *nummularii* (from the Latin *nummulus* for small money), they were the officers of the mint whose role was to test the quality of coins, that is, detect fake coins that were flooding the market. The *nummularii* also competed with the *argentarii* in offering banking services. The

American economist and economic historian Peter Temin (2004) argues that Rome's developed financial structure, whose level of sophistication he equated to that of modern advanced agrarian societies, had the potential to promote economic growth.

The Roman banking system, with financial transactions based on credit, suffered a number of crises due to excessive lending. Felix Martin (2013) relates how, when the Emperor Tiberius's officials were convinced that a recent boom in private lending had become excessive, they decided to reinforce an old law on capital adequacy instituted by Julius Caesar. The resultant credit crunch forced the Imperial Treasury to issue a three-year, interest-free loan program worth 100 million sesterces. To put this sum in perspective, estimates of Roman GDP at the time vary between 10 and 20 billion sesterces.

In Europe, the barbarian incursions into the Roman Empire which started in earnest in the 4th century AD produced the collapse of the economy, what the archeologist-historian Bryan Ward-Perkins (2006) called a collapse of the material culture and the end of complexity. When in the late Middle Ages there was a rebirth of economies and of exchange in the West, the use of credit instruments enabled economic development. The Florentine Bardi and Peruzzi merchant-banking companies became the most important banking houses in Europe but collapsed in the mid-14th century, a collapse traditionally attributed to King Edward III of England's repudiation of 1.5 million gold florins of war loans.

With the collapse of the Bardi and Peruzzi banking companies, the Medici and successively the Fugger family in Augsburg became the most important banking houses in Europe. Jakob Fugger "the Rich" lent large sums first to the Archduke Sigismund of Tyrol and later to the emperor Maximilian I but was wise enough to back the loans up with a guarantee – the right to exploit silver and copper mines in Hapsburg lands. As the story goes, when Maximilian defaulted, Fugger – unlike the Bardi and Peruzzi in a similar situation – rather than going bankrupt became the richest man in Europe. In the late Middle Ages and early Renaissance, while metal coins were still the preferred type of money for daily transactions and wages, banks were issuing letters of credit for long-distance trade across Europe.

The notions of credit and banking have changed over time. The Yapese, as mentioned earlier, carved notches on a stone to register credit and debt; the Ancient Sumerians inscribed wedge-shaped marks on clay tablets. For bankers in Ancient Rome, transactions were based on credit; payment in coins was limited to small amounts. The misadventures of medieval bankers, lending hard currency to their clients and in particular to sovereigns, have been mentioned. But the letters of credit issued for purposes of exchange by the big medieval merchant houses were almost never converted into coins: debts and credits were cleared at the fairs, in particular

the annual fairs of Champagne where Italian and Flemish merchants met to exchange goods, the balance of transactions typically being carried forward.

3.1.5 Paper money / banknotes

While paper money as banknotes appeared for the first time in China around the 7th century, it wasn't until 1661 that the first banknotes were issued in Europe, by the Bank of Sweden. It was a short-lived experiment: the Bank failed three years later. The first steady supply of banknotes started with the establishment of the Bank of England (BoE) in 1694. The BoE-issued banknotes were redeemable in gold until 1933, at which point the banknotes became legal tender: conversion to gold was no longer guaranteed. In what was to become the United States, banknotes were in circulation as early as 1775 but it was only in 1862 that the federal government declared the dollar the only state currency and only in 1913 that the Federal Reserve System was established. The backing of the U.S. dollar with gold was guaranteed until 1971.

3.1.6 Paper money becomes fiat money

Fiat money is a physical token, intrinsically worthless since the gold standard was abandoned, but with a nominal value, which is used as a medium of exchange and a store of value. Today's fiat money is declared legal tender by the issuing government. It is used to pay taxes and other sums due to the government and must be accepted in settlement of purchases or debts. It is the present stage of the monetary evolutionary process that began with commodity money, then adopted bullion and successively coins of precious metal, only to replace the bullion or coins with certificates convertible into the precious metal, and lastly issue certificates without the possibility of converting the certificates into precious metal, that is, fiat money.

As observed here, the use of commodity money is a form of barter where one good has the specific characteristics of being in short and relatively stable supply, divisible, and transformable by some technological process (e.g. melting) into something with intrinsic value. Fiat money, on the contrary, can be any object that can be used to purchase things in the market as legal tender but that has no intrinsic value and cannot be physically transformed into anything with intrinsic value. From the point of view of economic modelling, absence of intrinsic value means that the variable money cannot be placed inside a utility function. The sole use of money is to purchase things which have intrinsic value.

Fiat money can ultimately be thought of as credit money in that it represents purchasing power. It cannot be converted into anything at a predetermined rate but represents the ability of an individual to buy goods and services that are offered for sale. In this sense, fiat money is an anonymous measure of purchasing power.

Some economists, for example Wallace (1978), question whether fiat money ever existed before our time, arguing that, even if in several historical periods the convertibility of banknotes has been suspended due to, for example, emergencies such as war, in general convertibility was reinstated. Considering it unlikely that convertibility into gold or silver will be reinstated, Wallace concludes that today's banknotes are probably the first instance of fiat money to have ever existed.

The possibility of physically storing fiat money is essential for its use as storage of value. An essential characteristic of today's fiat money is its non-convertibility into gold or other. But the question of redeeming fiat money into metal is a bit tricky. Actually, in a sense, fiat money can be redeemed into anything that is offered for sale, including gold or other metals. But because it occurs at market price, the purchase of gold (or of anything else for that matter) does not qualify as "redeeming." The fundamental point is that there is no fixed exchange rate between fiat money and anything in the market.

Note that saying that the value of fiat money is not linked to the value of any physical "thing" is slightly misleading. Fiat money does not have value in itself but it is used to purchase things that have a price and thus value. In Chapter 6 we discuss how fiat money acquires value, that is, how prices are established in a given fiat-money setting.

Embodied in banknotes in most modern economies, fiat money is anonymous and not linked to any characteristic of the holder. Regardless of the financial situation of a person or entity, a cash transaction can always be made assuming that there is the required amount of banknotes. Bankruptcy laws might forbid the execution of some transactions with, for example, a firm in distress but this has nothing to do with fiat money in itself. With credit money, on the other hand, one needs credit to execute a transaction.

An important question with a bearing on the modelling of fiat money (see Chapter 4 for models of fiat money) is whether, in aggregate, there is a financial positive value made of cash. In the case of commodity money, money has an intrinsic value and can count among the components of global wealth. If banknotes are redeemable into gold or silver then we can also say that, in aggregate, there is financial wealth: it is the gold or silver that guarantees the pool of banknotes in circulation. Redeemable banknotes are a claim on some stock of precious metal held by banks and governments or central banks. In the case of fiat money, however, there is no physical

wealth associated with fiat money. At any moment there is a stock of banknotes of a given currency in circulation. Each banknote has a nominal value which allows the holder to buy things whose price is equal to the nominal value of the banknote. Persons who hold banknotes have financial wealth in that they can buy things like food, homes, cars, etc. Cash is part of the wealth of an individual after deducting debts.

But is this true in aggregate? Do banknotes in circulation count as wealth? The answer is somewhat conventional and depends on how we define terms such as "wealth." If fiat money is issued by a central bank, printed banknotes appear as liabilities of the central bank once they have been distributed. In modern economies, central banks give banknotes to banks against reserves, that is, the banknotes are sold or lent to commercial banks by their respective central banks. In both cases, on a central bank's balance sheet assets and liabilities sum to zero. In this sense there is no global aggregate wealth. Central banks might also give banknotes to purchase some asset(s); central banks' policy of quantitative easing is an instance of this. In this case also, the sum of assets and liabilities is zero.

The fact that banknotes are issued by an entity with a balance sheet is, however, a technical device that puts constraints on the distribution of money. The balance sheet constraints of the central bank preclude the free distribution of newly printed banknotes by banks to selected entities. In fact, as mentioned earlier, banknotes are exchanged for assets or lent to banks. They are distributed by banks to individuals or entities with a bank account. The sum of bank accounts plus banknotes is approximately equal to the net value of assets that the central bank has purchased or freely distributed (actually there might be deviations from this amount if loans are not repaid and need to be cancelled as credit). If we consider that currency in circulation and central bank accounts are liabilities of the central bank, again, there is no financial wealth in aggregate. This discussion on financial wealth in aggregate is based on the constraints to the distribution of money and is explained in, among others, Wray (2012). Wray argues that for every financial asset there is a corresponding financial liability so that the total sum is zero.

Suppose now that banknotes were created by some government printing press and distributed with some rules, for example uniformly to all citizens, but that they are still considered fiat money. There would be no balance sheet constraints; banknotes would not be a liability of any entity. Can we say that there is financial wealth in aggregate? The answer is again somewhat conventional. As we have seen, operationally, money is a tool to transfer ownership and to plan activities such as production. In aggregate, the only thing that can be done with fiat money is to transfer the ownership of physical wealth among agents. These transfers might, of course,

have an effect on the price of things so that the global value of things might go up or down. But we cannot consider the change in the value of things as a change in the value of money.

3.1.7 Bank deposits / credit money

After discussing commodity money, coins, and fiat money, let's now consider the form of money prevalent in modern economies: bank deposits. Bank deposits are credit money. As we will discuss in Chapter 5, money as deposits is generated by the private banking system, with the role of the state limited to (1) imposing, through base money, a unit of account; and (2) controlling interest rates. In modern financial systems, credit money can be defined as follows:

> *Credit money is a bank account, that is, a number written in the records of a bank (now typically a computer record) that represents the purchasing power of the owner of the bank account. This purchasing power can be used to make payments and can be converted into base money (coins and banknotes).*

Convertibility of deposits into base money is currently guaranteed but it is not essential for the purpose of conducting business such as payments including to the state. Payments can be handled through checks, bank transfers, or third-party electronic payment systems such as PayPal. Note, however, that should the elimination of banknotes which today serve as base money become an actuality, some other common unit of account would have to be established.

The notion of credit money needs to be carefully defined as it can be misleading for two reasons: (1) the term "credit" has more than one meaning and (2) the notion of credit applies to both the nature of money and its distribution. As for the different meanings, the term credit is generally used in three ways. First, it appears in a credit–debt relationship. If A owes something to B, we say that A has a debt and B has a credit. The nature of the debt can be money or other. For example, if we borrow sugar from a neighbor the nature of our debt is sugar; if we borrow money from a friend the nature of our debt is money; a trader who is going short of a certain stock has a debt of that stock.

Second, the term credit is, however, also used to indicate, generically, creditworthiness, that is, the right to enter into a credit/debt relationship. For example, the holder of a credit card has the right to buy things up to the limit of credit of his or her credit card. In this case the cardholder simply has a right to acquire a certain amount of goods and services and pay later. Stated differently, cardholders are allowed to enter into credit/debt relationships up to the limits of their credit. This

is different from the third notion of credit as purchasing power. Owners of bank deposits have credit as purchasing power in the sense that they transfer their purchasing power in payment of goods and services with checks or bank transfers. We should therefore distinguish between a credit relationship between two economic agents A and B where A owes something specific to B, and credit as a record of one's purchasing power or creditworthiness. The three types of credit are transferable but are conceptually different.

In modern economies, bank deposits are both a measure of purchasing power and a true credit. In fact, bank deposits can be used to buy goods or services and pay with checks or bank transfers. But bank deposits are also redeemable into banknotes or coins, either on demand or with some delay. However, the convertibility of deposits into banknotes is a marginal aspect of financial transactions and might be eliminated in some future time.

These two aspects of credit associated with bank deposits should not be confused with the fact that in modern financial systems, in aggregate, new bank deposits are allocated to those who take out loans. We will follow a common usage in the literature and call bank deposits credit money. However, this notion is not precise as credit money can signify simultaneously that the owners of deposits have purchasing power, that they can convert their deposits into base money, and that they got their deposits by taking out loans. In the literature the different meanings of credit are often confused.

Debt itself has a long history in human societies. Graeber (2011) argues that the notion of debt is primitive, and possibly related to slavery. He suggests that the first commercial transactions were essentially based on systems of credit and debt. We have already mentioned the example from ancient Sumeria where clay tablets recorded the credit–debt relationship between individuals and the temple. It is interesting to note that this relationship (i.e. creditor–debtor) could be transferred to other individuals.

Transferable credits were an integral part of the banking system in the Roman Empire and letters of credit were widely used as a safe instrument of long-distance trade across medieval Europe: a cloth merchant in Florence could buy a letter of credit from a big trading house paying in local currencies and then use the letter of credit to buy goods – say, raw English wool – in another location without the need to transport gold or silver coins.

But it was only with the creation of the first central banks that money as credit made its debut. The Bank of England started operations in 1694 to solve the financial problems of the Crown. It did so by issuing banknotes in exchange for loans redeemable into gold. In the U.S. at the time of the Civil War and in the absence of a

central bank, it was private banks that issued credit notes that circulated as money, again redeemable, at least in theory, into dollars or gold. Today money in the form of bank deposits is the vast majority of money in circulation as can be seen in Exhibit 3.1 which plots the amount of currency, banknotes, and total deposits in the U.S. economy for the 58-year period 1959 to 2017.

The theory of money as credit is not new: Schumpeter dates it back to Plato; the theory of money as a precious metal is older still – it goes back to Plato's teacher Aristotle. Schumpeter himself thought that money is credit in the sense of social accounting. In the 19th century, the view that money is credit was energetically proposed against the view that money must have some intrinsic value by the editor of *The Economist* Walter Bagehot. Bagehot (1873) observed that exchange in the business world is based on credit, not on the exchange of something intrinsically valuable such as gold. He argued that the view that money should have intrinsic value runs contrary to business practices and, if enforced, would be detrimental to the economy.

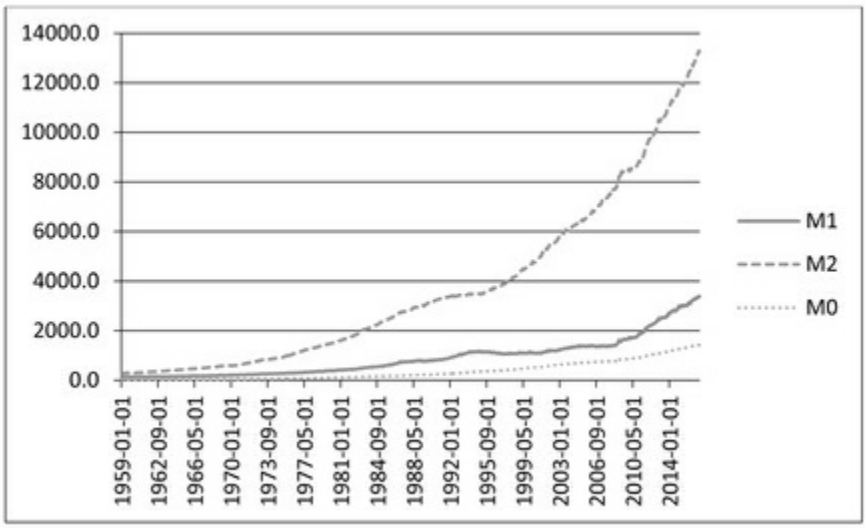

EXHIBIT 3.1 The evolution of the currency components M0, M1, and M2 in the U.S. economy for the period January 1, 1959 to January 1, 2017. *The currency component M0 includes coins and banknotes; M1 includes coins and banknotes as well as travelers' checks, and demand deposits; M2 includes M1 plus savings and time deposits and balances in retail money market and mutual funds. Constructed by the author using data obtained from Federal Reserve Bank of Saint Louis.*

The first modern formulation of the theory of money as credit is generally attributed to the Scottish economist Henry Dunning Macleod (1889). The starting point of Macleod's analysis of money is the observation that economic exchanges are not perfectly symmetrical insofar as we buy and sell things of different value at different times. In so doing we acquire credits and debts. For example, a farmer might sell a cow, acquiring a credit that allows him/her to buy items such as corn or

tools later. The farmer is a creditor but there is no specific counterparty as debtor. The essential character of credit money is the existence of a scoreboard that records the amount of credit of each individual.

Macleod wrote:

> This universally exchangeable merchandise is termed Money: and these circumstances show its fundamental nature. Its function is to represent the Debts which arise from unequal exchanges among men . . . and to enable persons who have rendered any sort of services to others, and have received no equivalent from them, to preserve a record of these services, and of their Rights to obtain an equivalent satisfaction from some one else, when they require it.
>
> *(p. 83)*

He then concluded: "The true nature of Money is now apparent. It is simply a Right or Title to demand some Product or Service from someone else . . . the true nature of Money is merely a Right or Title to acquire some satisfaction from some one else, i.e., a Credit" (p. 83).

The theory of money as credit was expanded and reformulated by Alfred Mitchell-Innes. Mitchell-Innes was a distinguished British diplomat, not a professional economist, but he had brilliant economic intuition. His first paper "What is Money?" (1913) received a favorable comment from Keynes, which encouraged him to develop his thesis in a second paper, "The Credit Theory of Money" (1914). In his 1913 paper, Mitchell-Innes defines the precise meaning of credit and money:

> Credit is the purchasing power so often mentioned in economic works as being one of the principal attributes of money, and, as I shall try to show, credit and credit alone is money. Credit and not gold or silver is the one property which all men seek, the acquisition of which is the aim and object of all commerce.
>
> *(p. 394)*

Credits and debts are formed in the practice of commerce: "The constant creation of credits and debts, and their extinction by being cancelled against one another, forms the whole mechanism of commerce . . . Credit and debt have nothing and never have had anything to do with gold and silver" (p. 395).

A year later, Mitchell-Innes (1914) discussed the theory of money as credit, explaining how a sale and purchase is the exchange of a commodity for credit. Therefore the value of money does not depend on the value of any metal or metals, but on the right to payment.

Both Macleod and Mitchell-Innes discuss credit money as credit that an individual acquires for selling something or rendering some service, thereby linking credit money to the notion that an individual (or entity) generates money with his or her activity. But this is misleading. Money as a bank deposit does not inform on *how* it was generated. A bank account is a purely abstract quantity that represents

numerically our ability to purchase goods or services, without informing on its source.

As noted earlier, three concepts of money as credit need to be reconciled:

- money as transferable credit, in the sense that bank deposits can be converted into banknotes;
- money as transferable purchasing power;
- money as deposit originated from a loan.

Most discussions on money as credit use the notion of money as transferable purchasing power. As mentioned earlier, Macleod himself considered money to be transferable debt or credit. In most economies today, bank deposits are credits in the sense that clients are entitled to receive, on demand or with other temporal arrangements, the value of their deposit in banknotes. In this sense the notion of money as credit depends on a more fundamental form of money; today that more fundamental form of money is banknotes. These credits can be transferred by, for example, writing a check or making an electronic transfer.

Neither Macleod nor Mitchell-Innes define credit as deposits convertible into banknotes or coins; rather it is clear from the above that both define money as purchasing power. It is this purchasing power that gets transferred to other individuals. Money is transferable credit in the sense that it is transferable purchasing power via checks or bank transfers. The ability to convert deposits into banknotes and coins loses its relative importance as we move towards a cashless society. In the writings of Mitchell-Innes, the notion of money as credit becomes the theoretical statement that money can always be identified with credit even if we use gold coins. As Keynes later famously wrote, coins are banknotes written on metal. Thus the essence of money is to be a record of purchasing power.

Credit has two essential prerequisites. The first is trust. For a system of credit money to work, each economic agent, that is, each individual or entity must trust that the bank functions properly and that their credit with the bank will be accepted in payment by other economic agents. Martin Shubik (2010) writes, “A universally accepted money is a perfect surrogate for individual trust. Thus information and reputation are dominant factors. The other properties are commentary” (p. 3). Except perhaps when coins have the same intrinsic value and face value, in all other cases, whatever practical instrument we accept as money we do so because we trust that others will do the same.

Trust is an essential component of any theory of money and the way money is allocated plays an important role in building trust. Again Shubik (2010) writes:

> A combination of technological and logical requirements for the efficient functioning of an exchange economy call for the use of an easily identifiable, simple, generally accepted means of exchange. Lack of trust calls either for a commodity money whose intrinsic value serves as its own guarantee; or, if fiat money is used, for laws and customs which are sufficiently enforced and accepted by the society and its government that the money is accepted as a societal surrogate for trust or value.
>
> *(p. 6)*

Second, for credit money to work, we need a safe system of transfer of deposits from one bank to another. Technically speaking, many systems for clearing the balance of credits/debts between banks are possible; in practice, all modern economies use a super-entity, a central bank, which ensures the proper functioning of transfers between banks.

The possibility of converting credit into banknotes is not essential but eliminating it would require, among other things, a mechanism to establish the value of deposits. In the current view, banks have the obligation to hand over – on demand – to their clients the equivalent in banknotes of their deposits. By the nature of today's banking business, this is very difficult if not impossible as the amount of deposits in the banking system largely exceeds the available banknotes. The Bank of England estimated that the value of outstanding banknotes in 2011 was only 11% of the total of outstanding credit in the country. In general, in conditions of normal trust of the banking system, depositors do not require cash. Actually in many countries today large cash transactions are forbidden by law. However, if depositors begin to suspect that a bank is not solvent, they might rush to take out their money in the form of banknotes, a situation known as a bank run. One of the functions of a central bank is to maintain the credibility of the banking system, standing ready to lend any bank enough banknotes to meet demand in the case of a run on the bank.

An eventual (total) suppression of banknotes would make it impossible to convert bank deposits into hard cash, thereby eliminating the possibility of bank runs and of bank failures. It is easy to imagine the psychological impact that a total absence of banknotes would have.

In practice, most newly created money today is associated with a loan or with asset purchases by central banks. This fact is empirically true as demonstrated in Exhibits 7.4–7.7 in Chapter 7. The entire mass of money in existence is backed by a parallel amount of either loans or reserves.

In a cashless system, banks would no longer have the obligation to deliver banknotes on demand but would have the obligation to make payments, that is, to transfer money, to other accounts on a client's demand. This means simply that an account is debited and another account is credited. The system works if every economic entity can purchase goods and services by debiting and crediting accounts.

Consider that deposits at banks are backed by loans. Clients who take out a loan

have the obligation to pay it back. Any increase in the total amount of deposits can occur only if new loans of an equivalent amount are taken out.

From an engineering point of view, the banking system is a huge scoreboard that assigns to each economic entity – individual, firm, or other – the amount of money the entity can use to acquire goods and services. The requirement that money be backed by loans is one way to regulate the allocation of money and induce trust in the monetary system. In principle a government could operate a scoreboard with money assigned to economic entities. But, of course, such an organization would have the problem of deciding to whom money should be allocated. The ideas of helicopter money to kick-start the economy or of a basic national income are examples of the creation of new money not backed by any loan, whose rule of distribution is simply uniform distribution.

3.2 Alternative forms of money

For a number of reasons – financial innovation, technological innovation, or the desire to create an alternative to legal tender to, for example, kick-start a local economy – various forms of money have been created and, it would appear, the pace of creation is increasing. Among the alternative forms of money that we will now discuss are what is called near-money and shadow banking, digital currencies, either private or created by central banks, and lastly local currencies. The latter are actually not so new; many date back to the Great Depression of the 20th century. While banknotes and credit money are easy to produce with modern technology, near-money calls for the use of complex financial engineering techniques; digital currencies required advances in cryptographic technology and distributed information systems.

Box 3.1 Aggregates of Money

Types of money available today include coins, banknotes, reserves, bank deposits, and alternative currencies including digital currencies. In determining their policy as regards interest rates – their main tool for managing the economy – central banks consider aggregates of money, that is to say, quantities fundamental for macroeconomic analysis. Among the aggregates are M0, M1, M2, MZM, and M4, with definitions that might differ from central bank to central bank.

For example, the money aggregate M1 includes coins and banknotes while M2 typically includes coins, banknotes, and bank deposits. But other aggregates can be considered. The Federal Reserve Bank of Saint Louis gives the following definitions for M1, M2, and MZM:

> M1 includes funds that are readily accessible for spending. M1 consists of: (1) currency outside the U.S. Treasury, Federal Reserve Banks, and the vaults of depository institutions; (2) traveler's checks of nonbank issuers; (3) demand deposits; and (4) other checkable deposits (OCDs), which consist primarily of negotiable order of withdrawal (NOW) accounts at depository institutions and credit union share draft accounts.
>
> M2 includes a broader set of financial assets held principally by households. M2 consists of M1 plus: (1) savings deposits (which include money market deposit accounts, or MMDAs); (2) small-denomination time deposits (time deposits in amounts of less than $100,000); and (3) balances in retail money market mutual funds (MMMFs).
>
> MZM (money with zero maturity) is the broadest component and consists of the supply of financial assets redeemable at par on demand: notes and coins in circulation, traveler's checks (non-bank

issuers), demand deposits, other checkable deposits, savings deposits, and all money market funds.

The Bank of England considers an additional aggregate, M4, whose definition is the following: "M4 includes notes and coins, deposits, certificates of deposit, repos and securities with a maturity of less than five years held by the non-bank private sector."

Similar definitions are adopted by the European Central Bank (ECB) and most other central banks in the developed countries.

For a full description of aggregates used by the Bank of England, see McLeay, Radia, and Thomas (2014, p. 10). For a full description of aggregates used by the U.S. Federal Reserve, see the Federal Reserve of Saint Louis website where they post their definitions of aggregates of money.

3.2.1 Near-money and the shadow banking system

"Shadow banking" is a term coined by Paul McCulley (2007) to indicate a system in which non-banks that are not regulated in the same way as deposit-taking institutions play a role similar to that of conventional banks, providing liquidity (money) to the market. Arguing that the shadow banking system provides fundamental services of market finance, Pozsar et al. (2013) write:

> Shadow banking activities consist of credit, maturity, and liquidity transformation that take place without direct and explicit access to public sources of liquidity or credit backstops. These activities are conducted by specialized financial intermediaries called shadow banks, which are bound together along an intermediation chain known as the shadow banking system.
>
> *(p. 7)*

Pozsar (2014, p. 8) defines the basic institutions and types of money as follows:

> There are four core institutions engaged in the issuance of money claims in the modern financial ecosystem: the central bank, banks (small and large), dealer banks and money market funds. These institutions issue four core types of money claims. The central bank issues reserves. Banks issue deposits. Dealer banks issue repos Money funds issue constant net asset value (NAV) shares.

Repos (short for Sale and Repurchase Agreements) are a type of debt contract by which a dealer sells a security S at time t at price P_S with the agreement to repurchase it at price $P_S + \Delta P$ at a later time $t + \Delta t$. The ratio $\Delta P/P$ is called the repo rate. Money market funds are funds formed with safe, liquid instruments that keep a constant value and can be sold at par on demand.

Repos, money market funds, and other very liquid assets are referred to as "near-money." But can we consider near-money to be true money? Does the shadow

banking system create money? Pozsar (2014) hints that near-money is a different type of money. Money as bank deposits is a costly tool borrowed to make transactions; it is thus not a good storage of value. To store financial value in a very liquid form, firms and individuals therefore use near-money. While near-money cannot be used to make payments directly it can be converted at par into bank deposits which can be used to make payments. So while near-money does not create purchasing power, it *stores* purchasing power.

Michell (2016) suggests that there is a kind of symbiotic relationship between money and near-money. Money creates purchasing power but it is not suitable as a store of value; near-money is a good, liquid store of value but does not create purchasing power. For this reason, they can be considered symbiotic.

3.2.2 Digital currencies

Let's now discuss digital (or crypto) currencies. Cryptocurrencies are based on blockchain – a cryptographic technology, [9] hence the name cryptocurrencies. While these currencies are typically referred to as "coins" or "tokens," digital currencies are not coins at all: they are simply an electronic means of payment that can be used to buy goods and services.

We will discuss separately private digital currencies already in the market such as Bitcoin, Ethereum, and Zcash, and the eventuality of digital currencies issued by governments or central banks. Our discussion of digital currencies starts by introducing the enabling technology: the distributed ledger.

Distributed ledgers / Blockchain technology

Money, as often said in this book, is purchasing power and any purchase involves the transfer of funds. The Bank of England's Financial Infrastructure Directorate looked at the connection between the emergence of digital technologies and payment technologies. Ali et al. (2014) write:

> Money and payment systems are intrinsically linked. In order for an asset to function as a medium of exchange, there needs to be a secure way of transferring that asset – a payment system. And for any system other than the exchange of physical banknotes or coins, a means of recording the values stored is also needed – a ledger.
>
> *(p. 1)*

Traditional payment systems are based on a two-tiered system at the heart of which there is a central ledger and a central authority – typically the central bank –

across whose books settlements take place. Commercial banks hold an account (reserves) with the central bank that is reflected in the central ledger of the central bank but also in the ledger of each participating bank; transfers of money between holders of bank accounts are done through the movement of reserves. The system is based on trust that transactions will be correctly implemented.

While the attention of the media and the public has focused on specific digital currency builders, such as Bitcoin, Ethereum, and Zcash – essentially marketing issues – the important innovation behind digital currencies is the distributed ledger, enabled by blockchain technology. Blockchain technology is what allows the shared (open) record-keeping and processing system to secure the transfer of the digital money between individual users without the need for banks or central banks while impeding the copying and reuse of the currency. The transaction, however, must be validated. Bitcoin transactions, for example, are validated in a decentralized approach by independent "miners" [10] estimated to number upwards of 5,000. There is no longer the need for a central authority. Ali et al. (2014) identified three innovations necessary for the distributed ledger: (1) cryptology to secure communication, (2) game theory for strategic decision-making, and (3) peer-to-peer networking without central coordination.

Private digital currencies

Private digital currencies are generated alongside the state-issued currency by certificates "bought" over the Internet with legal tender or exchanged on digital currency exchanges such as Coinbase, bitMEX, and Bitstamp. The innovation consists in the settlement technology, that is, the distributed ledger known as blockchain discussed earlier.

Cryptocurrencies can be thought of as securitized chunks of data (computer files) on a hard drive that use the web as a giant virtual safe; their introduction was enabled by increased access to the Internet and advances in computer and encryption technologies. The first private digital currency (the bitcoin) was introduced in 2009 by Bitcoin. The person or group who founded the currency did so under the name Satoshi Nakamoto; eight years since, Nakamoto's identity remains unknown. It is estimated that since the bitcoin's introduction, more than 840 blockchain-based currencies have been built.

Because cryptocurrencies are not backed by an issuing authority and have no intrinsic value – their value being determined only by demand – their builders try to control the generation of their currency, making it scarce. For example, by design, a

maximum of 21 million bitcoins can be created; an estimated 16.5 million bitcoins have been created to date. But precisely because cryptocurrencies have no intrinsic value, their downside is exposure to exchange rate fluctuations. Consider the bitcoin/U.S. dollar exchange rate. After remaining relatively stable at under or about $10 for the first four years after its inception, by end-2013 a bitcoin traded for $979.45 only to trade at $214.08 just over a year later (January 2015), then reached a peak of $3,500 at the beginning of August 2017 and was trading above $8,000 by the end of November 2017. Clearly holders of cryptocurrencies might enjoy large gains or suffer large losses.[11]

The cryptocurrency most widely used presently is the bitcoin. End-October 2016 it was reported that there were over 350,000 bitcoin users and almost 16.5 million bitcoins in circulation. Between the currency's inception in 2009 and end-October 2016, an estimated 241,528,102 bitcoin transactions took place;[12] the cost of a transfer today is about 83 U.S. cents. Advantages of the currency include ease of transfer and anonymity as the owner of bitcoins is obscured, though anonymity is not total: the blockchain ledger being open, the flow of funds can be analyzed.

Electronic coins (or "utility tokens") are now being created and "sold" by startup technology firms that want to raise money – an alternative to going to venture capital firms for funding. It is estimated that in the first half of 2017 alone, initial coin offerings (ICOs) have allowed startups to raise the equivalent of $1 billion. The tokens can be exchanged for other types of tokens on digital currency exchanges known as crypto-exchanges. An example is Bittrex. As this system is conceived now, crypto-coins are sold to potential investors.[13] The sale uses a hierarchy of money where new crypto-coins are sold in exchange for state-issued money or better-established crypto-coins.

Do cryptocurrencies qualify as money? As already observed, money is hierarchical. What can and what cannot be considered money is somewhat arbitrary. This means that there is some latitude in defining what we consider money. Based on our discussion of shadow banks, there are three conditions that must be met for an asset to qualify as money: (1) liquidity, (2) convertibility at par with legal tender, and (3) "acceptedness" as a means of payment.

Based on the above criteria, it is questionable if digital currencies qualify as money. They are "manufactured," sold, and can be used to purchase goods and services from suppliers who accept them; the latter includes the travel company Expedia and the U.S. online retailer Overstock.com. However, digital currencies are not very liquid, do not convert at par, and their acceptability as a means of payment is limited. It would be somewhat puzzling to consider digital currencies as money while near-money does not have the status of money. However, these considerations

are subject to change in function of the evolution of digital currencies.

State digital currencies

Private digital currencies are a kind of virtual commodity used in exchanges. The sheer amount of private digital currencies in circulation, the instability of their value, and new applications including their role in raising funds for technology startups (ICOs) have governments and central banks taking note. China's Fan Yifei, vice governor of the People's Bank of China and head of the central bank's cryptocurrency research, is behind the bank's push to supervise private digital currencies.

Besides supervising private digital currencies, a key question for the People's Bank of China (and governments and central banks in general) is whether to develop the central bank's own digital legal tender. Among the advantages: replace physical cash to cut the cost of issuing and circulating traditional money in the form of coins and banknotes. Though the People's Bank of China doesn't have a timetable for the launch of its digital currency (or at least has not yet disclosed it), Fan foresees that, once created, the bank's digital legal tender would circulate together with cash for quite a long time before replacing cash.[14]

There are important aspects that need to be considered in replacing traditional money with state-issued digital currencies. Digital currencies as currently provided by private entities are scarce by design and trade at market value. If issued under state control, digital currencies could be manufactured in variable quantities and trade at par with legal tender.

The People's Bank of China is not the only central bank considering the possibility of introducing a crypto version of fiat currency, in what has been qualified as a "game changer." The Bank of England's deputy governor Ben Broadbent (2016) discussed the possibility of households holding digital accounts directly in the central bank via software, bypassing commercial banks. Nevertheless, experiments with interledger technology at both the Bank of Canada and the Bank of England have concluded that the technology presently underpinning digital currencies – distributed ledger technology (DLT), the shared database and cryptographic system – is not yet sufficiently mature.

Carolyn Wilkins, senior deputy governor of the Bank of Canada and Gerry Gaetz president of Payments Canada, evaluated their joint experimental wholesale interbank payment system, using DLT (Wilkins and Gaetz 2017). Their conclusion: while it would be possible to settle wholesale payments on a distributed ledger,

there were presently problems with scalability and the confidentiality required around transactions in wholesale payments systems. Nevertheless, the Bank of Canada together with partners will continue to experiment, eventually exploring types of DLT-based financial market infrastructures, domestically or internationally, with the ability to settle multiple asset classes such as bonds, or money market instruments discussed above on the same ledger. More will be said on this in Chapter 7 where we discuss questions related to the distribution of money.

3.2.3 Local (alternative) currencies

The term *local alternative currencies* is applied to currencies with geographically limited circulation. Some of these currencies are, or were, created following the ideas of Silvio Gesell, a widely travelled merchant, political activist, and economist born in 1862 in what is now Belgium but was, at the time of his birth, Prussia. Gesell introduced the idea of *free currency*. His main idea was that money should be used for transactions, not for savings and not for speculation. He therefore argued that cash should earn no interest and, if not used, should lose value progressively. Gesell was ignored by mainstream economists of his time, and is still poorly known today though Irving Fisher[15] and John Maynard Keynes found his work of interest.

We will discuss four of the hundreds of local currencies that have been created: the German WÄRA, the Austrian Wörgl, the Swiss WIR Franc, and the America Ithaca's Hours. Each of these currencies represents the different notions and ideas behind local currencies.

The German WÄRA and the Austrian Wörgl were created around 1930 at the time of the Great Depression and remained in circulation only a few years; the Swiss WIR was created at the same time but still circulates today; the Ithaca's Hours were created in 1991 and still exist.

The WÄRA was created as Germany entered into a deep economic recession along with the rest of the Western world. A coal mine in Schwanenkirchen (Bavaria), which provided work for the region, had closed down due to the bankruptcy of the owners. The engineer Max Hebecker acquired the bankrupt mine at an auction with the intention to restart operations but lack of financing blocked his project. Inspired by the ideas of Gesell, Hebecker, himself a free-money activist, persuaded a free-money organization in Berlin to lend him 50,000 WÄRA, the currency that the organization had created. The WÄRA loan allowed Hebecker to reopen the mines and pay the miners. Local shops and artisans as well as the local government were persuaded to accept WÄRA as payment. To ensure circulation of the currency and

avoid its hoarding, Hebecker implemented the ideas of Gesell, stipulating that every month WÄRA would lose 1% of its value. This was achieved with the simple expedient of requiring that WÄRA were stamped every month and that stamps cost 1% the value of the currency.

As the WÄRA could circulate only in the region and lost value if saved employees spent all the WÄRA they earned – in the region. As a result there was an immediate resurgence of the local economy and Schwanenkirchen recovered its prosperity. The fame of the Schwanenkirchen experiment spread but soon met the hostility of the German central bank and of the central government. The currency was declared illegal; the region sank back into impoverishment.

A similar experiment with a similar result occurred in the small town of Wörgl, in Austria. In the 1930s the town suffered from the Great Depression, with unemployment and poverty rampant. In 1931, the recently elected mayor Michael Unter-guggenberger decided to tackle the problem by issuing a local currency following the ideas of Gesell. Unterguggenberger used the little finance available to the town as a guarantee and issued stamp scrips, a money circulating only locally and whose value was lost over time by the expedient of the need to stamp the scrips monthly. As in Bavaria, the experiment proved successful, gaining recognition as far away as the U.S. where it came to the attention of the economist, statistician, and inventor Irving Fisher. Tens of other towns in Austria were interested in implementing a local currency but, as in Germany, the experiment met the hostility of the Austrian central bank. The stamp scrips were declared illegal and their production stopped in 1933.

Unterguggenberger was denied permission to enter Switzerland and address conferences but that did not stop an initiative by the Basle-based cooperative, the WIR Bank, to create its own alternative (interest-)free money, the WIR Franc, a about the same time. Similar to other ideas inspired by Gesell, the non-interest-bearing WIR Franc was stamped monthly to lose value if stored. The WIR Ban successively abandoned the principle of stamping the currency to discourage its hoarding as well as the idea of interest-free money. The WIR Franc – now an electronic currency – is used by 62,000 Swiss small- and medium-sized enterprises.

A more recent attempt to create a local currency comes from the small New York State town of Ithaca – home of Cornell University. Ithaca's Hours were created in 1991 by Paul Glover as a tool to favor local business and for social progress. A former professor at Temple University and community organizer, Glover's objective was to create a local currency whose nominal value is approximately equivalent to an hour of work (thus the currency's name Ithaca Hours) that would circulate locally, thereby fostering local business and local cohesion. Though still in circulation, Ithaca

Hours have been hurt by the widespread use of electronic means of payment.

There are several similarities in the above attempts to create alternative local currencies. First, these initiatives have as their objective to foster local economic development. The local nature of the alternative currency and the fact that it loses value if not used produced a high velocity of circulation and boosted local business activity. The idea of using newly created money to put people to work and their salaries to produce more work by spending is typical of Keynesian thinking. In the 1930s, when most alternative currencies were generated, a major problem (outside of unemployment itself) was the hoarding of money which reduced the amount of money in circulation and consequently consumption. Methods to reduce the hoarding of money were one way to boost economic activity. Today, the situation has changed: money is credit. In aggregate, consumers do not hoard money but borrow it to maintain their standard of living. To store money, people now tend to invest in treasury bonds or stocks or other financial assets.

As mentioned, a key ingredient of these local alternative currencies is the limited diffusion within a geographically well-defined group with multiple mutual exchanges, that is, a small (relatively closed) economy. In this they differ from cryptocurrencies which operate globally across the Internet and whose success as a medium of exchange is dependent on a large, diversified community of users.

The historical account of the evolution of money from barter is reflected in many mainstream theoretical models of money, where money is essentially a tool to overcome the inefficiencies of barter. We will discuss these models in Chapter 4, but first let's look at the competing concepts of money.

3.3 So just what is money? Metallists and Chartalists

While, as we have seen, there are many forms of money, economists would like to find some commonality, some basic unifying concept. Recall that we are now considering market economies based on the notion of private property and that money is what allows the free transfer of private property and the storage of value for future use. Money is a social mechanism, that is, something strictly linked to institutions, that allows not only exchange and storage but also the organization of complex economic activities, such as industrial production. Money is what allows the coordination of the free efforts of individuals or entities towards a common objective. In the absence of money, coordination is obtained through command.

Two competing concepts of money have been put forward: Metallists and Chartalists. This dichotomy has been cast in different forms. Charles Goodhar (1998) observes that Metallists are those who maintain that money must have an intrinsic value or at least be backed by an intrinsic value while Chartalists[16] argue that money is created by an issuing authority, typically a State, and has value because the State declares it has value and that it can be used to pay sums due including taxes. In the Metallist camp we can list early thinkers including the philosophers Aristotle and, much later, Locke, and the economists Jevons, Menger, Von Mises, Alchian, Kiyotaki, and Wright. In the Chartalist camp, we can place the philosopher Plato (teacher of Aristotle) and, much later, the economists Knapp, Mireaux, Graziani, Wray, Keen, and Goodhart himself.

Goodhart further observes that there is a parallel dichotomy between those economists who believe that money is a product of the private sector which optimized the barter process and Chartalists who hold that money was created by the State. The idea that money originated in the private sector is due to Carl Menger (1892), founder of the Austrian School of Economics. But there are other characterizations of Metallists and Chartalists. Metallists are often identified with those economists who believe that money is something that must be "manufactured" while Chartalists, especially Circuitists,[17] argue that money is essentially credit. (Recall the different notions of "credit" associated with money.) This last characterization of money by the Metallists would lead to their acceptance of manufactured digital currencies such as the bitcoin as money while the previously mentioned definitions of both the Metallists and Chartalists would not recognize digital currencies as money.

A more abstract notion of money as credit or, more precisely, as purchasing power was proposed and has gained acceptance. When we sell something including our

labor in exchange for money, we acquire a credit to purchase things in the future. Thus money is identified with purchasing power. The concept of money as credit in the sense of purchasing power is both an empirical fact and a conceptual generalization. Note that the notion that money is credit, a recompense for actions such as selling goods or performing services, is slightly misleading. How one gets a bank deposit is a question of the allocation of the deposits, not of its source. For example, people might inherit bank deposits without having performed any service that entitles them to the deposit.

Box 3.2 Tomorrow's Money

A number of possible changes in the nature of money are now under discussion. The first change concerns the elimination of coins and banknotes. Coins are nearly irrelevant from the standpoint of a theory of money; even banknotes now represent only a small fraction of the entire stock of money. Coins and banknotes are anonymous and therefore transactions based on either are not traceable. Some governments would like to eliminate both coins and banknotes and replace them with bank deposits and electronic means of payment which are traceable. In November 2016, in a demonetization experiment whose objective is to curb black money and corruption, the Indian government outlawed 86% of the currency in circulation. The full impact of the "cash ban" is not yet clear.

Still coins and banknotes represent the base money issued by governments while bank deposits are created by banks. In modern banking systems, banknotes play no essential role in creating bank deposits. However, banknotes which are issued by the state represent the numeraire of money. As bank deposits are convertible into banknotes, their value is the value of the banknotes into which they are convertible.

If banknotes were suppressed, bank deposits would need to be numerically anchored to some other base measure of value. Depending on the measure, this might put the generation of money in the hands of the private sector; the value of bank deposits would become a private convention. Suppose that two or more banks want to adopt different conventions as regards the value of their deposits. There must be an explicit conversion rate between these values. But because governments must specify the values at which they accept payment of taxes and other levies, governments must find a way to fix a common measure of value.

Were coins and banknotes to be eliminated and bank deposits not convertible

into any base currency, a bank deposit would not be a debt of the bank in the usual sense of the term debt. In fact, without convertibility, a bank does not owe anything to its depositors, but would only be obliged to transfer any given portion of the deposit to other banks and to accept transfers from other banks. If bank deposits are granted after according a loan, then the obligation of extinguishing the loan becomes the obligation of having, at a given date, the requisite amount in the bank account to allow for the loan to be extinguished and the bank account reduced by a corresponding amount.

But even in the above examples, money remains the indicator of the purchasing power of individuals or firms. Today this purchasing power is borrowed from banks in return for interest payments and acceptance on the part of the borrower to return the sum plus interest at some future date. Money is the certificate of deposit. Convertibility into a currency allows one to fix the value of bank deposits; loans allow the distribution of new money. However, once we have an understanding of the nature of money it is clear that rules regarding its use and distribution can be modified.

The introduction of digital currencies, if espoused by governments and central banks, might change the game in some fundamental way. Today digital currencies are a financial epiphenomenon based on a private entity selling an electronic commodity made scarce by design. The adoption of digital currencies on a large scale would imply solving a number of "engineering" challenges, and among them are the following:

1. The distribution or allocation of digital currencies (electronic certificates) as purchasing power is a social issue: the rules of distribution need to be established and accepted independent of the technology;
2. How to control eventual large-scale digital currencies raises serious questions on (1) accessibility: does everyone have access to a computer and the Internet? and (2) security: does everyone have confidence/trust in the system?;
3. It has to be decided if the absence of traceability of money transfers is socially and legally acceptable as well as its contrary, i.e. if it is acceptable that no transfer of funds escapes government surveillance;
4. Technological innovation might quickly render any technology obsolete: legal tender based on encryption techniques, blockchain, and the Internet must be capable of adjusting to even radical changes in technology, and sustain the costs involved.

Historically, money as credit was an explicit relationship between a creditor and a debtor. Referring again to the Sumerian example, their temples had an elaborate system of tablets that recorded the debt and credit not only of temple employees but also of farmers and artisans. Debt and credit constituted a well-defined relationship of individuals with the temple. However, the notion that all money is credit, as formulated by Mitchell-Innes, is a purely conceptual generalization. It is essentially the statement that any form of money implies that the holder of money has "credit," that is, the ability to purchase goods or services. This almost philosophical generalization has, however, an important practical consequence: money does not need to have intrinsic value and, even if it does have intrinsic value, what is really important is the amount of purchasing power that money gives its holder.

A similar interpretation of the function of money was proposed by Joseph Schumpeter (1934). In Schumpeter's view, though money is essentially credit, the fundamental function of money is to serve as social accounting. Schumpeter stressed the logical nature of money in both the Metallists' and Chartalists' theories.

The debate on the nature of money is not a truly scientific debate on money but more an engineering question as well as a debate on economic policies. Money is a human artifact and its nature is defined by the rules and institutions that were designed to support it. In addition, tradition plays a role and the terms used in defining money inherit meaning from the past.

Notes

1 This example was taken from Gordon Childe (1942).

2 The author wishes to thank Carlo Angiolini, who grew up in Piancastagnaio during and after World War II for this information.

3 The source is Robert Bartlett (1993, pp. 282–283).

4 King Alyattes of Lydia (now part of modern-day Turkey) is credited with having minted the first official currency, a standardized coinage used in long-distance trade in the Mediterranean.

5 Note that the role of gold or silver in Ingham's "military-coinage complex" is not limited to the Ancient World. In writing about the post-Roman world, Chris Wickham (2009) cited a major strength of the Ottonians: the silver mines discovered in their Saxon heartland near Goslar at around 970 AD. The mines funded the kings – and above all their large army – for two centuries (p. 434).

6 Using Greenland ice evidence of hemispheric lead pollution, Hong et al. (1994) estimated that world production surpassed Roman levels only in the middle of the 18th century.

7 "Seigniorage" comes from Old French and refers to the right of the feudal lord or seigneur to mint money. Today seigniorage is used to signify the difference between the face value of the money and the cost of producing it. If the seigniorage is positive, the issuing entity makes a profit; if it is negative, the issuing entity makes a loss.

8 See Guy Fourquin (1990, p. 43).

9 Cryptography or cryptology is the theory and practice of encoding messages so that they can be read and understood only by selected individuals who have the keys to the encoded message. While diplomats and secret agents have always used cryptographic techniques, these techniques are now being used to encode messages related to payments. Modern cryptography is based on the disciplines of mathematics, computer science, and electrical engineering.

10 Though miners are said to be independent, most miners belong to a mining pool. More will be said on miners in Chapter 5 where the creation of digital currencies is discussed and, in particular, the role of "miners" in verifying transactions.

11 Source: CoinDesk: www.coindesk.com/price/

12 Source: BNC: www.quandl.com/collections/markets/bitcoin-data.

13 In July 2017, the U.S. Securities and Exchange Commission (SEC) – an independent federal government agency responsible for protecting investors and maintaining the fair and orderly functioning of securities markets – decided that tokens issued through the ICOs might be considered securities, thereby subjecting them to regulatory scrutiny, including disclosure requirements.

14 Source: Bloomberg News. 2016. "Central Banks Should Lead on Digital Currency, PBOC's Fan Say

(September 1).

15 Referring to Silvio Gesell's theory of free money, Yale University professor of economics Irving Fisher had this to say: "Free money may turn out to be the best regulator of the velocity of circulation of money, which is the most confusing element in the stabilization of the price level. Applied correctly it could in fact haul us out of the crisis in a few weeks . . . I am a humble servant of the merchant Gesell." Source:*Stamp Scrip*, New York: 1933, p. 67; and *Mail and Empire, Toronto*: November 21, 1932. Fisher's research on the Quantity Theory of Money (discussed in Chapter 8) was to lead to monetarism, a school of thought that stresses governments' role in controlling the amount of money in circulation.

16 Goodhart uses the spelling Cartalists but in the literature Chartalists is perhaps more frequently used.

17 The Circuit Theory of money is a heterodox theory of money creation. The Circuit Theory of money argues that money is endogenously created by banks, used to pay wages, and then destroyed. It emphasizes the essential distinction between banks and industrial firms. According to Circuit Theory, credit is the primary form of money. For a discussion of the Circuit Theory of money, see Chapter 8.

References

Ali, Robleh, John Barrdear, Roger Clews, and James Southgate. 2014. "Innovations in Payment Technologies and the Emergence of Digital Currencies." *Bank of England Bulletin 2014 Q3*. www.bankofengland.co.uk/publications/Documents/quarterlybulletin/2014/qb14q

Bagehot, Walter. 1873. *Lombard Street or a Description of the Money Market*. London: Kegan Paul, Trench, Trübner and Co.

Bartlett, Robert. 1993. *The Making of Europe: Conquest, Colonization and Cultural Change 950–1350*. London: Allen Lane.

Broadbent, Ben. 2016. "Central Banks and Digital Currencies." Speech given at the London School of Economics. (March 2): www.bankofengland.co.uk/publications/Pages/speeches/2016/886.aspx.

Childe, Gordon. 1942. *What Happened in History*. Harmondsworth, UK: Penguin Books.

Davies, Glynn. 1996. *A History of Money*. Cardiff: University of Wales Press.

Fourquin, Guy. 1990. *Histoire Economique de l'Occident Médiéval*. Paris: Armand Colin:40–44.

Goodhart, Charles. 1998. "The Two Concepts of Money: Implications for the Analysis of Optimal Currency Areas." *European Journal of Political Economy*, vol. 14, no. 3:407–432.

Graeber, David. 2011. *Debt: The First 5,000 Years*. New York: First Melville House.

Hong, Sungmin, Jean-Pierre Candelone, Clair C. Patterson, and Claude F. Boutron. 1994. "Greenland Ice Evidence of Hemispheric Lead Pollution Two Millennia Ago by Greek and Roman Civilizations." *Science,* New Series, vol. 265, no. 5180 (September 23): 1841–1843.

Ingham, Geoffrey. 2004. *The Nature of Money*. Malden, MA: Polity Press.

Jevons, Stanley. 1875. *Money and the Mechanism of Exchange*. New York: Appleton and Co.

Macleod, Henry Dunning. 1889. *The Theory of Credit*. London: Longmans, Green, and Co.

Martin, Felix. 2013. *Money: The Unauthorized Biography – from Coinage to Cryptocurrencies*. London: The Bodley Head/Random House Group.

McCulley, Paul. 2007. "Teton Reflections." Global Central Bank Focus 2, September, PIMCO.

McLeay, Michael, Amar Radia, and Ryland Thomas. 2014. "Money Creation in the

Modern Economy." *Quarterly Bulletin* (Q1). Monetary Analysis Directorate, Bank of England.
Menger, Carl. 1892. "On the Origins of Money." *Economic Journal* 2:239–255; translation is by C.A. Foley.
Michell, Jo. 2016. "Do Shadow Banks Create Money? 'Financialisation' and the Monetary Circuit." PKSG Working Paper 1605.
Mitchell-Innes, Alfred. 1913. "What is Money?" *The Banking Law Journal*, vol. 30 (May): 377–408.
Mitchell-Innes, Alfred. 1914. "The Credit Theory of Money." *The Banking Law Journal*, vol. 31 (December/January):151–168.
Pozsar, Z., T. Adrian, A. Ashcraft, and H. Boesky. 2013. "Shadow Banking." Economic Policy Review, Federal Reserve Samuelson Bank of New York.
Pozsar, Zoltan. 2014. "Shadow Banking: The Money View." Working Paper 14-04, Office of Financial Research. (July 2).
Samuelson, Paul. 1976. *Economics*. London: McGraw-Hill, International Student Edition.
Schumpeter, Joseph. 1934. *The Theory of Economic Development: An Inquiry into Profits, Capital, Credit, Interest, and the Business Cycle*. Cambridge, MA: Harvard University Press.
Shubik, Martin. 2010. *Theory of Money and Financial Institutions*. Cambridge, MA: MIT Press, vol. 3.
Temin, Peter. 2004. "Financial Intermediation in the Early Roman Empire." *The Journal of Economic History*, vol. 64, no. 3 (September):705–733.
Wallace, Neil. 1978. "The Overlapping Generations Model of Fiat Money." Staff Report 37. Federal Reserve Bank of Minneapolis.
Ward-Perkins, Bryan. 2006. *The Fall of Rome and the End of Civilization*. Oxford, UK: Oxford University Press.
Wickham, Chris. 2009. *The Inheritance of Rome: A History of Europe from 400 to 1000*. London: Allen Lane.
Wilkins, Carolyn and Gerry Gaetz. 2017. "Could DLT Underpin an Entire Wholesale Payment System?" (Contributed to) *The Globe and Mail*. (May 25): https://beta.theglobeandmail.com/report-on-business/rob-commentary/could-dlt-underpin-an-entire-wholesale-payment-system/article35106771/
Wray, L. Randall. 2012. 2nd ed. 2015. *Modern Money Theory*. London: Palgrave Macmillan.

4
Modelling Money

When Renaissance architects wanted to sell their projects to their patrons, they presented wooden models of the building which allowed the patron to modify the project to obtain the desired result – in terms of status[1] and comfort if a villa or palace or of functionality if a church, the headquarters of a commercial enterprise, or that of an administration. Today, economists model money and the economy to allow governments and central banks to interpret and explain empirical facts and to gain a better understanding of the dynamics of money and the implications of monetary policies under consideration. The mathematical models of commodity money and banknotes that we discuss in this chapter were designed to be integrated with General Equilibrium Theories (GETs; see Box 4.1). Note, however, that the modelling challenge has changed: today's modelling challenge is to integrate a theory of money with non-equilibrium macroeconomic theory in a stock-flow consistent framework. The concept of stock-flow consistency (double bookkeeping in corporate finance) was introduced in Chapter 2; stock-flow consistent models are discussed in Chapter 8.

Box 4.1 General Equilibrium Theories

General Equilibrium Theories (GETs) are economic theories in which the economic system is formed by the utility optimization of rational agents under constraints of equilibrium. Perhaps the simplest way to illustrate GET is by presenting the Capital Asset Pricing Model (CAPM) of Sharpe–Lintner–Mossin (see Sharpe 1964, Lintner 1965, and Mossin 1966). The CAPM is an asset pricing model that determines the theoretical prices of assets. It is the simplest GET. To understand the logic of CAPM we have first to recall the rule of optimal investment stated by Harry Markowitz, co-recipient of the 1990 Nobel Memorial Prize in Economic Sciences. In laying the foundation of modern portfolio theory, Markowitz considered the problem of an investor who has to choose a portfolio of N investments. Assuming that the investor knows the vector of expected returns $\mu = [\mu 1 \ . \cdot \cdot \ \mu N]'$ and the variance-covariance matrix of returns Σ, Markowitz argued that a rational investor should choose a portfolio whose weights $w = [w_1 \ldots w_N]', \sum_{i=1}^{N} w_i = 1$ (weights are the percentages of each

asset in the portfolio) maximize the following expression: $\boldsymbol{\mu}'\omega-\lambda\omega'\sum\omega$. This expression is the expected return of the portfolio minus a penalty term given by the variance of the portfolio multiplied by a risk aversion parameter λ.

Based on Markowitz's portfolio theory, Sharpe (also co-recipient of the 1990 Nobel Memorial Prize in Economic Sciences), Lintner, and Mossin created the CAPM. The CAPM assumes that all investors know the vector of expected returns, the covariance matrix of returns, and the risk-free rate and behave as prescribed by the Markowitz model. Under these assumptions, it can be demonstrated that all investors share the same risky portfolio which coincides with the market portfolio. The general equilibrium nature of the model consists in the fact that the weights chosen by each investor are in equilibrium with the weights of the naturally occurring market portfolio.

GETs are generalizations of this basic idea. First, GETs consider not one single period but a finite or infinite period. Quantities such as prices as well as all the other information of models are not simple random variables but stochastic processes in continuous time. Agents have rational expectations in the sense that they know the characteristics of the stochastic processes used for optimizing. The object to be optimized is the expectation of a utility function defined on the stochastic processes.

4.1 Modelling coins

Let's now discuss the modelling of coins. You might ask what's the purpose of modelling coins, a form of money of little relevance in today's advanced economies. The answer is: models of coins typically address problems which remain pertinent today. In particular, models of commodity money try to solve the following problems:

1. *The stability of prices under different regimes*: Metallists typically argue that a monetary system based on money with intrinsic value is more stable than a monetary system based on credit or fiat money. A number of models address the problem of stability quantitatively.
2. *The quantity of coins to be minted under free minting*: With free minting, individuals can decide the proportion of gold or silver to mint and the proportion to keep in a form with an intrinsic utility, such as jewelry (we will use the word "jewelry" as it is generally understood in texts to indicate all uses of gold or silver other than coinage). Models address this issue as a utility optimization problem.
3. *The behavior of bimetallic systems*: A bimetallic system, where the ratio of the value of gold and silver coins is fixed, is exposed to the problem that the ratio of the market values of silver and gold might be different from the legal ratio. This might cause a bimetallic system to become de facto a monometallic system if one of the coins disappears.
4. *The efficiency of different systems*: A system based on commodity money subtracts precious commodities from their productive use. Models address the quantitative comparison of systems strictly based on commodities with systems at least partially based on fiat money.

Consider the question of introducing money in the general equilibrium models that have been the mainstream macroeconomic equilibrium models since the 1980s. In these models, there is no variable and no process that represents money. In neoclassical economics, it is held that money has little or no effect on real-world economic processes. In the framework of an equilibrium model, it is difficult to explain why agents hold money – a commodity considered to have no utility. But despite this view, it was felt necessary to introduce money in equilibrium models.

The seminal papers on modelling money are Robert Lucas (1982), Lars Svens-son (1985), and Lucas and Nancy Stokey (1987). Lucas and Svensson both adopt the idea of cash-in-advance constraints (CIA Constraints) originally proposed in Rober

Clower (1967). The CIA constraint is simply the condition that, in any period, an agent must have enough cash to purchase the desired goods or services. The constraint of having enough cash proved to be a fruitful modelling idea. Lucas and Svensson assume that the government has an exclusive right to produce money which is in turn distributed through fiscal policies. Lucas and Stokey (1987) extended the model by introducing a choice between cash and credit.

Let's now introduce several models of coins. Observe that each model adopts a specific view on how coins are minted, by whom, and to whom they are distributed. We start with two models, the one due to Thomas Sargent[2] and Neil Wallace and the other due to Thomas Sargent and Bruce Smith.

The Sargent and Wallace (1983) model of commodity money is an overlapping-generation model where agents live through two periods. The model includes a consumption good (generically called "bread") and a capital good (generically called "gold"). Bread represents anything that has a utility for consumers, gold represents money that is needed for transactions but does not produce any utility. The model assumes that there is a mechanism to transform bread into gold and vice versa. Though at first sight this assumption might seem strange, it simply reflects the fact that gold can exist in two different forms: coins which do not bear any utility; or gold as a metal for making "jewelry," which has utility but can also be a consumption good.

The model is set in discrete time. At time 1 the economy receives the initial endowment of gold. No other endowment of gold is received. However, at each time t the economy receives a random amount of bread. It is possible to convert time t bread into time t gold and to convert time t bread into time t+1 gold but it is not possible to convert time t bread into time t+1 bread.

Agents have a twice differentiable, quasi-convex utility function. The model is cast in the form of an equilibrium model that maximizes agents' utility. Several different versions of this model can be constructed, including models with multiple investment goods.

For an idea of the effects of the transformation of metal from the state of jewelry to the state of coins, consider the example of the Viking raids during the period from roughly 800 to 1000 AD. Whether the raids were provoked by the expansion of the Franks or other, the wealthy monasteries in England and elsewhere where gold and silver were "stored" in ecclesiastical objects (bread) were a prime objective. The Vikings minted this wealth into coins and transformed the economy of northern Europe from an economy based on ecclesiastical wealth into a mercantile market economy, with a trade network that was to cover all of modern Europe, Russia, the Middle East, northern India, and China. Coin hoards of 9th-century dirhams (the currency of the Arab caliphs) discovered in Scandinavia demonstrate the extent of

Viking trade with the Middle East along the Volga.[3]

Sargent and Smith (1997) propose a model of coinage, debasements, and Gresham's law which states that "bad money drives out good money" cast in the framework of free minting. The authors write:

> For hundreds of years, supplies of coins in Europe emerged, via a curious mechanism, from voluntary decisions of owners of old coins and bullion to exchange them at mints for new coins. Mints were sometimes private enterprises, licensed to produce on demand a list of coins whose design and fineness was specified by the sovereign. Citizens were free to take metal to the mint to purchase newly minted coins, often bearing less rare metal than was surrendered for them . . . A debasement was a respecification of the list of coins which could be produced, with the rare metal content (the fineness) of coins of each nominal value being reduced.
>
> *(p. 1)*

In the Sargent–Smith model of commodity money, coins circulate at face value, the *tale*. It is therefore possible that coins circulate with values higher or lower than the market value of their content in precious metal. The objective of the authors is to explain how it might happen that citizens take old coins to the mint to have them reminted with a lower content of precious metal. Another of Sargent and Smith's objectives is to model Gresham's law. The observation that bad money drives out good money had already been made by Nicole Oresme in 1350 and by Nikolaus Copernicus in 1519 but it was later attributed to Sir Thomas Gresham, an English financier who lived under the Tudor dynasty. "Bad money" is coins with a metal content whose market value is less than the nominal value of the coins, "good money" is formed by coins whose value is equal to the market value of their metal.

The global setting of the Sargent–Smith model is the following. There is a set of households that live infinitely. Time moves in discrete steps. There is a single consumption good and money in the form of silver coins. Coins can be carried from one period to another but do not yield utility. Consumption cannot be stored and carried to the next period. There is a mechanism to convert consumption into coins and vice versa.

At each time step each agent receives a random endowment of the consumption good. There is a CIA constraint to ensure that each agent has enough money to pay for what he/she wants to consume in the next period. The entire model works by optimizing the utility of agents in an equilibrium framework.

Velde and Weber (2000) proposed a model of bimetallism. The authors summarize the debate on bimetallism as follows: "Many issues raised in the debate fall under two broad questions. The first concerns existence: Is bimetallism implementable, sustainable, stable? The second question concerns welfare properties: Is bimetallism desirable?" and conclude: "The model formally confirms [that . . . c] oncurrent circulation is viable over long periods without changes in the legal ratio . . .

bimetallism does provide a more stable price level, [but . . .] a single standard always yields higher welfare than bimetallism" (p. 2).

Velde and Weber's model is cast in discrete time without an end. There are three goods: a consumption good c and stocks of gold and silver bullion. Cont sumption is non-storable while bullion can be stored and does not depreciate. There is an exogenous endowment $X_t = (\xi_t, Q_{1t}, Q_{2t})$, where ξ_t can be consumption subject to $ct \leq \xi_t$ and Q_{1t}, Q_{2t} are the stocks of gold and silver bullion. X_t is stochastic and its probability distribution is known to agents.

Bullion exists in two forms: metal which is a source of utility to consumers, and coins. By metal is intended any form that is not coin, for example "jewelry." The percentages of bullion in metal are d_{1t}, d_{2t}. Consumers obtain utility from consumption and from ownership of metal. Preferences are represented in the following form:

$$E_0 \sum_{t=0}^{\infty} \beta^t \left[u(c_t) + v(d_{1t} + d_{2t}) \right]$$

where u, v are twice-differentiable functions and $u'(0) = +\infty$. Coins, which are characterized by their weight b_i, are produced by a competitive firm which can transform metal into coins or vice versa at zero cost. However, a tax σ_i must be paid on each new coin.

Coins are needed to meet the CIA constraint. At the beginning of each period the endowment is known. The consumer can sell his/her endowment to the minting firm, thereby acquiring a claim on its profit and can buy consumption and metal at prices p_t, h_{it}, $i = 1, 2$.

A more recent model is that developed by Angela Redish (2011) at the Federal Reserve Bank of Minneapolis. The author describes her model as follows:

> We build a random matching model of monetary exchange in which agents need to use coins as a medium of exchange to buy (perishable) goods. We assume an exogenously fixed quantity of a non-perishable good (silver) that can be held in a utility-yielding form (which we call "jewelry") or can be held as coins that do not yield utility directly but can be traded with other agents. Agents can acquire coins (at a cost) by taking jewelry to the mint or can acquire jewelry by melting coins. The monetary authority determines the size of (amount of silver in) a coin and the number of types of coins in existence.
>
> *(p. 2)*

As in Velde and Weber's model, there are two components: a consumption good and silver. The consumption good cannot be stored. Silver can be held as "jewelry" or coins. Coins of two different weights are minted. Consumers obtain utility from jewelry and consumption but not from coins. Time is discrete and infinite. Each period of time is divided into two sub-periods. In the first sub-period agents are

randomly selected to be consumers or producers of the consumption good. After being selected they trade, that is, they exchange money for the consumption good. In the second sub-period, they can melt coins to obtain jewelry or mint jewelry to obtain coins. The value of coins is their weight in silver. The total amount of silver is kept fixed but the number and weight of coins are variable. Agents maximize utility. The dynamics of the model is given by utility maximization.

4.2 Modelling fiat money

Let's now discuss the modelling of fiat money, for which there are two main families of models: overlapping-generation (OLG) models and search-and-matching models. Models based on the CIA constraint or on dynamic general equilibrium models discussed earlier are suitable for commodity money but not for fiat money.

We start with OLG models. The idea of overlapping generations was first put forward by Maurice Allais[4] (1947) and successively by Paul Samuelson[5] (1958). The various versions of OLG models differ on some basic assumptions. A good starting point is the model in Wallace (1978) to which we will primarily make reference.

Suppose that time moves in discrete steps, from time 0 to infinity. Agents are all equal and live for two periods, in the first period as young, in the second period as old. At each step $N(t)$ young agents enter the economy, live one period as young, and then live the following $t + 1$ period as old. Lifetime is equal to two periods for all agents. At each step there are therefore $N(t) + N(t - 1)$ agents. In the simplest model, the number of agents is constant equal to and therefore at each step there are agents. More complex OLG models make demographic assumptions such as $N(t) = nN(t - 1)$, $n > 0$.

The economy produces a single consumption good $c(t)$. Call $c^y(t)$, $c^o(t + 1)$ the consumption of the typical representative agent at times t, $t + 1$ respectively. All agents have the same consumption. The representative agent has a utility function $u(c^y(t), c^o(t + 1))$when young and simply $c^o(t + 1)$ when old. That is, when old the agent simply maximizes consumption. Different models make different assumptions as regards the production and the characteristics of the consumption good. The simplest assumption is that each agent receives an amount $w^y(t)$ when young and an amount $w^o(t + 1)$ equal to zero when old. Other models allow old agents to receive a positive amount of consumption good $w^o(t + 1) > 0$. However, $w^o(t + 1)$ must be less than their optimal consumption in the old stage. In the simplest model, the consumption good is assumed to be perishable, so no saving is possible. In other systems, the consumption good can be stored, but at a cost.

To understand the role of fiat money, let's first look at the simplest system with zero endowment to old agents and perishable consumption. In this system, old agents simply die as they have no access to consumption. If we introduce fiat money, agents can sell consumption when young, get money, and buy consumption when old. To whom do young agents sell consumption? To the old of the previous generation.

The system works starting from an initial time with an old generation that was

endowed with a stock of money. In this initial time step, the young get a stock of the consumption good, selling a portion of their consumption endowment to the old who had been endowed with money at a previous step. Respecting appropriate constraints, the OLG economic system is set in motion. At time step 1 a new generation of young is endowed with consumption and the old from the previous step have money to buy consumption from the young. In other OLG systems, the consumption good is also perishable and cannot be stored, but old agents receive some consumption goods. In this situation, an equilibrium state can be reached even without money. Now introduce fiat money. This money can be stored at zero cost and can be put aside to save efficiently from one period to another. With money we can reach a better optimum, giving up some consumption when young to enjoy more consumption when old.

In still other OLG systems, the consumption good can be stored at a cost. As fiat money can be stored at zero cost, the introduction of money still allows a better optimum under appropriate constraints. Other extensions of the basic model include a periodic endowment of money in addition to the initial endowment. The basic structure and results remain the same: money allows for a better transfer of wealth from one generation to the next.

In the language of neoclassical economics, overlapping generations introduce basic "frictions" in a model of consumption. Money has no value in itself but becomes valuable because it overcomes the frictions of the overlapping generations and therefore allows optimization. OLG models allow one to use the familiar utility optimization framework with a good – money – which has no intrinsic value and which therefore cannot be placed as an argument of a utility function.

The value of money is the inverse of the unit price of the consumption good, that is, it is the quantity of money that can be acquired with a unit of the consumption good. Money has value if its supply is limited. In each OLG model, the value of money depends on its supply, which fixes prices and therefore the value of money.

In addition, in OLG models, money is valuable only if it has value in every moment. If the value of money goes to zero in one moment, then the value of money is zero in every moment. This feature of the OLG models translates the common notion that money has no intrinsic value but acquires value because we trust that other agents will accept its usage. In other words, in trading with an agent, we exchange goods for money because we have confidence that in the future we will be able to trade in the money for goods. Young agents give up consumption for money because they are confident that they will be able to exchange money for consumption goods when old. It is easy to see that if this condition fails at any one time, it fails every time. OLG models are intrinsically infinite time models. If the model stops at any finite time,

money has no value.

OLG models can reach multiple equilibrium states. The equilibrium of OLG models is called a *tenuous equilibrium* because it is based on expectations, not on fundamentals. Price paths are said to follow a *bubble path* because prices depend at each instant on future expectations. In a sense, OLG models are self-fulfilling prophecies.

Obviously OLG models do not faithfully reflect the real world. They cannot capture many of the phenomena typical of modern financial economics, including the fundamental disequilibria that are found in modern economies and the formation of growing inequalities between agents. In addition, the notion of the value of money depends on aggregating goods into a single composite good. In modern complex economies, this is typically impossible.

Let's now turn our attention to the other set of models that have been proposed to describe fiat money, the random search-and-matching models. The random search-and-matching models were first proposed by Peter Diamond (1984) and by Nobuhiro Kiyotaki and Randall Wright (1989, 1993).

The basic search-and-match theoretical model of money assumes that there are a large number of infinitely lived agents in search of a match. There are also a large number of consumption goods that come in a basic unit form, that is, the quantity of consumption goods can be only 0 or 1. The economy is endowed with an amount of money M. Each agent can either own a unit of money or produce a unit of consumption goods. After producing the unit of consumption good j, agent i searches for another agent with whom to trade.

It is generally assumed that agents prefer to not consume what they produce; the model therefore assumes sets of matching probabilities. If two agents find a profitable match, they trade and reenter the system. Agents might exchange a commodity for another commodity, money for a commodity, or a commodity for money. Encounters between agents are assumed to follow a Poisson process.[6] Money solves a barter problem: after randomly meeting, agents might not be able to barter (exchange commodity for commodity) while exchange with money can always take place.

We have described three main types of models of fiat money: models based on the cash-in-advance principle for modelling commodity money, overlapping generation models, and, lastly, random search-and-matching models. These models describe the emergence of money as an optimization process to overcome the limitations of barter. But, as discussed in Section 3.1.1, anthropologists now question the fact that commodity money or fiat money emerged in societies where exchange was based on barter; some argue that purely barter societies never existed but that the dominant

form of exchange was based on debt and credit.

Observe nevertheless that the absence of historical evidence of the emergence of money from barter is in itself not a reason to reject the models presented thus far. The problem with the models that we have discussed is the lack of empirical evidence in their favor; these models are oversimplifications of a complex reality. Money is a key factor responsible for disequilibria in economies. Equilibrium models of money can capture only general features of money but cannot capture some of the more important features of money in modern societies.

Notes

1 The Renaissance architect and builder of the dome of Florence's then new cathedral Santa Maria del Fior was asked by Cosimo de Medici, the richest banker of his time, to design a palace which would also serve as the headquarters of the Medici Bank. On seeing Brunelleschi's project, Cosimo refused it as too magnificent, saying, "Envy is a plant no man should ever water." A disappointed Brunelleschi smashed the model (Symonds 2017).

2 Sargent was a co-recipient of the 2011 Nobel Memorial Prize in Economic Sciences for his work on genera equilibrium models.

3 These coin hoards of dirhams were discovered on the island of Bjork (Birka in 9th-century written sources) by the 19th-century Scandinavian archeologist Hjalmar Stolpe who, over 20 years, excavated 1,100 buria mounds on the island. Source: Norse Mythology Research Project https://sites.google.com/site/norsemythologyresearchproject/researched-essay.

4 Maurice Allais was awarded the Nobel Memorial Prize in Economic Sciences in 1988 "for his pioneerin contributions to the theory of markets and efficient utilization of resources." Allais was professor of economics at the École Nationale Supérieur des Mines de Paris.

5 Paul Samuelson was awarded the Nobel Memorial Prize in Economic Sciences in 1970 "for the scientifi work through which he has developed static and dynamic economic theory and actively contributed to raising the level of analysis in economic science." Samuelson was professor of economics at the Massachusetts Institute of Technology.

6 A Poisson process is a model of events that occur at random times.

References

Allais, Maurice. 1947. *Economie et Intérêt*. Paris: Imprimerie Nationale.

Clower, Robert W. 1967. "A Reconsideration of the Microfoundations of Monetary Theory." *Western Economic Journal* 6:1–9.

Diamond, Peter. 1984. "Money in Search Equilibrium." *Econometrica*, vol. 52, no. 1 (January):1–20.

Kiyotaki, Nobuhiro and Randall Wright. 1989. "On Money as a Medium of Exchange." *The Journal of Political Economy*, vol. 97, no. 4 (August):927–954.

Kiyotaki, Nobuhiro and Randall Wright. 1993. "A Search-Theoretic Approach to Monetary Economics." *The American Economic Review*, vol. 83, no. 1 (March):63–77.

Lintner, John. 1965. "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets." *Review of Economics and Statistics*, vol. 47, no. 1:13–37.

Lucas, Robert E., Jr. 1982. "Interest Rates and Currency Prices in a Two-country World." *Journal of Monetary Economics* 10, no. 3:335–359.

Lucas, Robert E., Jr. and Nancy L. Stokey. 1987. "Money and Interest in a Cash-in-Advance Economy." *Econometrica*, vol. 55, no. 3 (May):491–513.

Mossin, Jan. 1966. "Equilibrium in a Capital Asset Market." *Econometrica*, vol. 34, no. 4:768–783.

Redish, Angela. 2011. "A Model of Commodity Money with Minting and Melting." Federal Reserve Bank of Minneapolis, Research Department Staff. Report 460. (July).

Samuelson, Paul. 1958. "An Exact Consumption-Loan Model of Interest with or without the Social Contrivance of Money." *Journal of Political Economy*, vol. 66, no. 6:467–482.

Sargent, Thomas and Neil Wallace. 1983. "A Model of Commodity Money." Federal Reserve Bank of Minneapolis, Research Department Staff Report.

Sargent, Thomas J. and Bruce D. Smith. 1997. "Coinage, Debasements, and Gresham's Laws." *Economic Theory,* vol. 10, no. 2 (August):197–226.

Sharpe, William F. 1964. "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk." *Journal of Finance*, vol. 19, no. 3:425–442.

Svensson, Lars E. O. 1985. "Money and Asset Prices in a Cash-in-Advance Economy." *Journal of Political Economy*, vol. 93, no. 5 (October):919–944.

Symonds, John Addington. 2017. *Renaissance in Italy: The Age of the Despots*. Norderstedt, Germany: hansebooks.

Velde, François R. and Warren E. Weber. 2000. “A Model of Bimetallism.” *Journal of Political Economy*, vol. 108, no. 6 (December):1210–1234.
Wallace, Neil. 1978. “The Overlapping Generations Model of Fiat Money.” Staff Report 37. Federal Reserve Bank of Minneapolis.

5
How Money is Created

In previous chapters we discussed the nature of money more as a question of engineering than as a truly scientific question: money is a human artifact that might change in function of the rules and institutions a society adopts. What we call a "theory of money" is in reality the understanding of how different systems create what we call money and its role in a given society. The definition of money is not unique as there is a hierarchy of assets that we might call money or near-money. Different institutional systems establish a different separation between money and near-money. This chapter discusses how the different types of money are created. Future chapters will discuss how money is allocated, how it gains acceptance and acquires value, and how its value changes.

5.1 The question of money generation

In itself, the creation of money is now a simple process which might, nonetheless, involve the use of sophisticated technology. Coins are minted with modern metallurgical processes; banknotes are printed with sophisticated printing processes to render counterfeiting more difficult. Creating bank deposits is also a simple process, essentially consisting of entering a number in a ledger, today invariably a computer file. But the process of creating near-money such as money market funds or the many other types of funds created by shadow banking systems can be more complicated and typically involves sophisticated financial engineering techniques.

While the creation of money is in itself a conceptually simple process, what are economically important are the conditions of money creation and its allocation, two interdependent processes. Today money creation is associated with (1) the central banks printing banknotes and engaging in buying assets in open market operations or in policies such as quantitative easing and (2) bank loans (i.e. credit money) extended by commercial banks. In the case of bank loans, the creation and allocation of money are part of the same process as new deposits are associated with those who take out loans. However, this association of loans/money creation did not hold in other periods and will not necessarily hold in the future.

Given the strict interdependence of the creation and allocation of credit money, we will discuss the two processes together. However, we will do so only for didactical reasons. Much of what textbooks consider the creation of money is conceptually the process of the allocation of money and this is because credit money cannot be created and stored awaiting its allocation but is jointly created and allocated: a bank deposit does not exist without an owner.

An important issue related to money generation is the hierarchical dependence of different forms of money and its evolution as discussed in Mehrling (2015). Hierarchical dependence means that the definition of and the existence of money of a given level depend on money of a higher level. Mehrling writes: “Always and everywhere, monetary systems are hierarchical. One way that economists have tried to get an analytical grip on this empirical fact is to distinguish *money* (the means of final settlement) from *credit* (a promise to pay money, or means of delaying final settlement)” (p. 394; italics in original).

Different institutional settings require different hierarchical dependence. For example, the issuing of banknotes initially depended on the existence of a more fundamental commodity money: in the case of Western economies, gold. As for bank deposits, money as bank deposits was initially related to banknotes. Banknotes are

the base money, deposits subordinated money. But this link has been relaxed as banks can now create loan–deposit pairs with only a weak link to banknotes. In (eventual) future cashless economies, banknotes might be abolished; the hierarchy of money would then start with deposits. When we consider money creation we have to keep in mind the creation of a hierarchical money system.

A key issue in the process of money creation is the impact of the quantity of money generated on the price level. Monetarists introduced simple relationships between the quantity of money and the level of prices. But, as will be discussed in Chapter 6, "price level" is not a well-defined, unique notion: there are different methods of defining an index of prices, and therefore there are different price levels. In addition, modern economies are segmented into multiple sub-economies and in many of these sub-economies, due to innovation and market effects, changes of the price level cannot be measured. As a consequence, it is not possible to define a single model of the relationship between the quantity of money and the price level. While the focus of our attention is on the creation of money in today's modern economies, we will also make a few brief remarks on the creation of other forms of money in the past.

The creation of money is not a uniform process. Actually there are coexisting processes of money generation that reflect the hierarchy of money. At the top is the creation of banknotes and reserves by central banks. Central banks (or the Treasury in some cases) can create money by minting coins, or printing banknotes. Central banks can also buy assets paying with banknotes they have printed or crediting, *ex nihilo*, reserves to the accounts of commercial banks. Central banks can also lend banknotes or reserves to commercial banks. Ultimately in most developed countries, central banks are the only organizations legally entitled to create money for their own use, for example to buy assets.

However, in modern financial systems where the largest part of the stock of money is credit money, money is created by commercial banks as paired loans/ deposits, as will be described shortly. Though in principle bank deposits can be converted into banknotes, banknotes are not essential for the creation of bank deposits. Here we find a slightly paradoxical situation where banknotes occupy the highest level in the hierarchy of money but, in practice, the public can obtain banknotes only by converting bank deposits into banknotes. There is no direct supply of banknotes to the public. Though central banks are not essential in the process of deposit generation, they have the essential roles of facilitating the transfer of funds between commercial banks, of guaranteeing the credibility of the system and, above all, of implementing fiscal/monetary policies such as setting interest rates.

One level below in the hierarchy of money are money market funds and other highly liquid assets that are created by the shadow banking system. These assets are

assembled as packages of other assets using sophisticated financial engineering techniques and are then sold in open markets primarily as stores of value.

Considering money generation from an engineering point of view, the problem facing the design of any system of money generation is clear: the system must be able to generate enough money to support transactions and eventually the storage of value, and must be socially acceptable from the point of view of money distribution. Money, which embodies economic power, cannot be distributed arbitrarily. The engineering solution today is a system that lends money for transactions while it sells near-money or other less liquid assets for storage of value. The purchase of assets and the consequent supply of money for transactions is considered an exceptional measure to be reversed in the near future. Note that here we clearly see the distinction between two concepts: (1) money as credit in the sense of purchasing power and (2) money supplied under a credit arrangement.

The present solution is a system based on three fundamental building blocks: (1) the central banks which provide base money as coins and banknotes and reserves as well as control functions, (2) the commercial banks which provide the bulk of money used in transactions, and (3) a shadow banking system which provides liquid assets for storing value, a function for which credit money is not well suited.

Banknotes, which are sold to the public by the commercial banks, provide the ultimate underpinning of value; credit money, which is allocated through loans, provides the tools for making transactions; and near-money, which is also sold, provides the storage of value. This system was partially engineered and partially evolved spontaneously under the pressure of market forces. This system is clearly in a state of evolution. Banknotes are losing importance, being progressively replaced by electronic means of payment, and might be abandoned in the future. Any eventual abandonment of banknotes would change the hierarchy of money and create the problem of establishing measures of value.

Let’s now turn our attention to the generation of different types of money. Our emphasis will be on the generation of credit money but we will briefly touch upon the generation of the various types of money in circulation today.

5.2 Creating traditional forms of money

Conceptually the generation of traditional money is a simple process even if it requires technology. For example, the use of bullion required a mastery of metallurgy; coins required in addition the use of dies to stamp the coins; the printing of banknotes was not feasible without the invention of an economical support such as paper and movable-printing devices. Let's now discuss the process of generation of traditional forms of money: commodities, coins, and banknotes.

5.2.1 Metals and other commodities as money

As observed by Charles Goodhart (1998), in the literature there is a distinction between money as a commodity with intrinsic value and money as a token without intrinsic value and a parallel distinction between money as something that is manufactured at a cost and money whose production cost is (near) zero.

The conceptual notion that commodity money is based on something rare, something in short supply, was a clever financial invention but it was a source of human suffering from the beginning. Metals (and other commodities) are not freely available. They have to be mined – sometimes at a huge human cost – bought, or otherwise procured. The Spanish colonial mint of Potosi (Bolivia), which produced an estimated 60% of all the silver mined worldwide in the second half of the 16th century, provides an example. Thousands of miners – native Indians and African slaves – lost their lives due to exhaustion, diseases such as pneumonia, and mercury poisoning. But already in Antiquity we have examples of the harsh conditions in the mines. In the north of Italy, the gold mines of La Bessa, near Biella (Piedmont) covering almost 10 square kilometers, are considered to be the largest open-cast gold mines in the world. It is estimated that thousands of slaves lost their lives there between the 2nd and 1st centuries BC when the mines were exploited by the Romans.

In his *Debt: The First 5,000 Years*, Graeber (2011) suggests that money followed on the commoditization of human beings in the form of slavery. He cites examples where the value of things was quoted in human beings, for example prices quoted in terms of Irish slave girls.[1] The Greek historian Diodorus Siculus (1st century BC), in his universal history *Bibliotheca*, gives us an idea of the exchange rate the Romans used in trading with the Gauls: an amphora of wine for a slave.

To avoid carrying too much weight, especially in long-distance trade, bulk commodities used as a medium of exchange had to be rare. Today we think of the quantity of money as a determinant of prices. If money is too abundant, prices go up.

The precise dynamic of this process depends on the system of reference that we adopt, that is, on how we determine a price level, but there is little doubt that creating more money produces an increase in prices, a process we call price inflation. The notion of price inflation has a parallel with the use of commodities as money. An abundant commodity has a low exchange value. Consider water. Nobody would use water as commodity money, at least not in countries where water is abundant. But a rare commodity which is in high demand has a high exchange value.

Now consider salt – a mineral that was needed to preserve food and supplement diets but which was stored in underground deposits not easily available, at least not to all peoples. Moorish merchants routinely traded salt gram for gram for gold in the sub-Sahara as early as the 6th century; Venetians were to pick up the trade and the exchange rate centuries later.

5.2.2 Minting coins

The creation of one of the simplest forms of money – coins – is not difficult to understand intuitively. The use of coins is ultimately related to the availability of (semi-)precious metals,[2] in particular gold or silver, and the need to standardize their weight and value. The metals might be obtained when sources are discovered and mined, via booty, exchange, or other means. Coins are created by the minting of (semi-)precious metals. The owner of the metal is the owner of the coins. Chapter 4 presented simple models of the amount of coins minted based on the optimization of utility between the "industrial" use of the metal, which has an intrinsic utility, and the monetary use of the metal, which has no intrinsic utility but serves to facilitate exchange.

Historically speaking, it is difficult to imagine that the minting of coins was done in function of the optimization of utility rather than in function of the needs of a "sovereign." Graeber (2011) observes that large armies needed a supporting market for servicing their needs, from food and shelter to prostitutes. Standing armies therefore created a market for the exchange of goods and services and, consequently, the need for money in some ready-to-exchange anonymous form.

As we saw in Chapter 3, anthropological studies in the West date the first minted coins to around 600 BC, when conquering Lydian armies took metal as booty and melted it down to pay soldiers' wages. Soldiers had to be paid and weapons forged in what Geoffrey Ingham (2004) called the "military-coinage complex"; it was not a question of the optimization of utility between the industrial use of the metal and the monetary use of the metal.

The use of commodities as money is a form of barter, common in ancient times when traders exchanged commodities such as amber or slaves for other commodities or manufactured goods such as wine and vases. Trading implied a primitive notion of exchange value. The minting of coins introduced a new discretionary element absent when one trades pure commodities: One needs to decide (i.e. optimize) what fraction of a given metal to mint. Gold and silver in the form of coins (as opposed to a pure commodity) have no added value but lubricate exchange.

Coins were conceived to provide metal of certified purity in standard formats: the value of a coin was the value of the metal of which it was made. More precisely, the metal of which coins were made was the unit of value. But by measuring the value of "things" in terms of the equivalent in metal, the value of things fluctuates in function of the relative value of the metal. If the metal becomes more or less rare or desirable, then the value of everything else changes. Changes in availability or desirability of a metal might have different sources, including new industrial uses of the metal. (The question of value will be discussed in Chapter 6.)

The situation is more complicated if coins are minted in two (bimetallism) or more metals. The problem is related to establishing the relative value of coins made of different metals: only one type of coin can have a value equal to the value of the weight of the metal. Bimetallism illustrates the situation in which coins might have a nominal value different from the value of the metal in which they are minted. If the value of the metal of which coins are made is higher than the nominal value of the coin, then coins with the higher content in metal tend to be withdrawn from circulation and melted.

To illustrate the above, suppose that there are both gold and silver coins, that both coins weigh 10 grams, and that the nominal value of the gold coin is ten times the nominal value of the silver coin. One could then exchange one gold coin for ten silver coins. Suppose, however, that in the market for metal, 10 grams of gold are worth 90 grams of silver. Then with one gold coin one can buy 90 grams of silver; alternatively, the same person can acquire ten silver coins, melt them, and be left with 100 grams of silver. Clearly silver coins tend to disappear.

Even with one metal, coins might have a nominal value different from the market value of the metal in which they were minted. The filing and clipping of metal (especially gold) coins are two of the most common ways to remove the metal from the coin. Metal can also be removed from the coins quite naturally by their circulation over time. When a coin's content in metal is inferior to its nominal weight, we say that coin has been debased. Debasement can also occur by edict. The sovereign or the government can issue an edict prescribing that the metal content of a coin is different from its nominal value.

The U.S. effectively abandoned the gold standard during the Great Depression (Great Britain had abandoned the pound's peg to bullion two years earlier in 1931) to allow the government to inject money into the economy and lower interest rates. The link was formally abandoned almost 40 years later, in 1971. It is estimated that to back the $2.7 trillion now in circulation and in deposits with the approximately 261 million ounces of gold held by the U.S. government, gold prices would have to rise to about $10,000 an ounce, up from its present market value of about $1,225 at the time of this writing.[3]

5.2.3 Printing banknotes

When we go from coins to fiat money such as banknotes and credit money such as bank deposits, the process of money creation becomes more complicated and, ultimately, more difficult to understand. In both cases, money no longer has intrinsic value and the cost of its production is close to zero. The amount of money created must therefore be determined in function of the consequences on the economy. In the case of fiat money, we discussed (see Chapter 4) simple overlapping-generation models: These models address the question of the quantity of money to create as the optimization of a barter process between two generations, the young and the old. In the case of both fiat money and credit money, the role of central banks is critical, as we will see later in this chapter.

Today most of us are used to paying small sums with coins and banknotes while the bulk of payments are made via credit cards, bank transfers, or other electronic payment systems. Despite the reduced role that banknotes play in modern economies, for many people banknotes still represent the embodiment of money and a wallet full of banknotes represents wealth. More importantly, in current banking systems, banknotes are the underpinning of credit money, and this is because bank deposits are convertible into coins or banknotes. This convertibility of bank deposits determines their value in terms of base money.

The nature of banknotes has changed over time. In the past banknotes appeared in two main forms: as certificates convertible into gold or as certificates convertible into some other currency. Today's banknotes are yet something else: fiat money not convertible into anything else if not into purchasing power.

The first banknotes appeared in China, the country that invented paper and then printing. Due to a shortage of metallic currency (copper coins), notes had already replaced coins as "burial money" in the 6th century – a common-sense instance of optimization. During the Tang Dynasty (618–907 AD), a familiarity with credit notes

used by long-distance merchants facilitated the acceptance of paper money and by the end of the Dynasty, merchants were depositing their metallic money in the government Treasury in exchange for official "compensation notes" called Fey-thsian or flying money.[4]

The concept of banknotes as certificates convertible into legal tender such as gold or silver coins was introduced in Europe in the 13th century by long-distance merchants such as Marco Polo, who reached Cathay (China). But it wasn't until four centuries later that banknotes appeared in Europe, first in Sweden (1661) and then in England, with the creation of the Bank of England (1694). These notes were quite different from the banknotes in circulation today: they were handwritten to a precise amount and issued on deposit or as a loan, with the name of the payee written on the note. Fixed denomination notes only appeared 50 years later. The advantage of the notes to sovereigns borrowing for war or short of precious metals was clear.

While at times in the past anyone who had gold could have it minted into coins, banknotes with full gold convertibility were typically emitted only by sovereigns: no other entity would have sufficient credibility to guarantee the full convertibility of the notes into precious metals such as gold. With full convertibility, the so-called gold standard, banknotes have the same value as gold. The sovereign's gold reserves (theoretically) limited the quantity of banknotes he could print.

In addition to sovereigns or governments creating banknotes as certificates convertible into gold, in the past, private banks were allowed to emit banknotes – essentially credit certificates. In the U.S. at the time of the Civil War, private banks extended credit to their clients and issued banknotes under a condition – that the value of the notes they issued did not exceed 90% of the value of the Federal bonds purchased by the bank and deposited in the Treasury, as stipulated in the National Bank Act (1863). These banknotes were debts issued by the private bank. Clients paid interest and eventually repaid the amount due, while the banknotes they received continued to circulate. If the banknotes were returned to the issuing bank for redemption, the bank was obliged to pay the bearer the amount of the banknote in gold. In the meantime, the bank had earned interest and (in principle) the principal was repaid. In cases when the banknotes never came back, the bank earned the full amount as profit.[5]

A private banking system of this type worked by creating a secondary system of banknotes as a medium of payment. The basic medium of payment remained gold coins and government-issued currency fully backed by gold. In addition, banks emitted banknotes that circulated as money. Banknotes guaranteed by private banks clearly had credibility that, though not perfect, was much higher than the credibility of the individual borrowers signing an IOU.

Progressively, the gold backing of banknotes was weakened. Reserves were not sufficient to cover the entire value of the banknotes printed by the issuing authority. Consider that according to the World Gold Council, about two thirds of all gold ever mined has been mined since 1950. Were all this gold melted and transformed into a cube, the cube would measure approximately 21 × 21 × 21 meters and it would weigh about 187,000 tons. Assuming a price of $40,000 per kilogram, the current value as metal of all gold ever mined is approximately:

$187,000 × 1,000 × 40,000 = $7,480,000,000,000, or $7,480 billion.

Compare this to the M2 U.S. money stock as per June 19, 2017, which is $13,510 billion. Though the U.S. currently holds more gold than any other country – approximately 8,133 tons of gold or roughly 4.3% of all the gold ever mined – clearly the value of all the gold ever mined would not be sufficient to cover the U.S. M2 money stock; the government would not be able to convert its currency into gold.[6] The actual value of all the gold the U.S. holds is $8,133 × 1,000 ×40,000 = $325 billion – against a total of $1,479 billion of coins and banknotes currently in circulation, that is M1.

Partial convertibility into gold marks a fundamental change in the nature of banknotes as money. With full convertibility banknotes are essentially replacement for coins; with partial convertibility banknotes become fiat money. Money circulates not because of its intrinsic value but because it is declared legal tender by the issuing authority.

The final step is to declare banknotes legal tender without any guarantee of convertibility into something with intrinsic value, such as gold. As mentioned previously, the Bank of England took this step in 1931; the U.S. did so 40 years later in 1971. Lack of convertibility into gold implies (1) that the price of gold can fluctuate, that is, gold is no longer the reference; and (2) the amount of banknotes in circulation is no longer linked to the amount of gold reserves.

For obvious reasons, the printing of banknotes has to be under strict government control. To discourage counterfeiting, the production of banknotes has progressively become more sophisticated. Japan's National Printing Bureau, which produces the Bank of Japan notes, listed some of the technologies they use to make it easier to identify counterfeit notes. The list includes intaglio printing, watermarking which produces variations in the thickness of the paper, holograms which when tilted produce various images, color-shifting inks, luminescent inks or pearl ink which produces a pink pattern at the center of the note when tilted, and microprinting. But in the past, when the sophisticated techniques that we now have did not exist, other methods were the main defense against counterfeiting. During the Song Dynasty in

China (960–1279), for example, the penalty for counterfeiting was the standard punishment: beheading.

5.2.4 Issuing letters of credit

Letters of credit – a "letter" guaranteeing that a buyer's (importer's) payment to a seller (exporter) will be made according to the terms and conditions stated in the written document – were first used in China around 1150 AD and by the 13th century, with the unification of Asia under the Mongols, their use spread first to the Islamic Empire and from the Islamic Empire to Western Europe via Italian merchants. Once introduced into Europe, letters of credit were widely used in long-distance trade where they were understandably preferred to bags full of gold coins.

A letter of credit from a major Florentine trading house was accepted in most trading places. On arrival in Champagne, for example, the merchant could purchase goods and eventually obtain local currencies. The recipient of the letter of credit was typically another major trading house, which in turn emitted letters of credit. Once or twice a year, the big trading houses met at major fairs such as the fairs of Champagne (France) and cleared their letters of credit. Balances were typically carried forward. Today letters of credit are bank guarantees of payment transferred electronically between banks.

5.3 The creation of bank deposits and the multiplier

The vast majority of modern money takes the form of bank deposits. Bank deposits are a convenient implementation of credit through the banking system transferable to virtually any economic agent. In the current mode of implementation, the creation and allocation of bank deposits are intertwined in the sense that new deposits are created when loans are subscribed or deposits of exogenous funds are made. Conceptually, however, the creation of deposits and the allocation of deposits are two separate processes.

The economist and politician Augusto Graziani remarked that money requires specialized firms for its creation. In modern systems these firms are the commercial banks. The banking system guarantees the easy transferability of deposits between agents and provides the trust needed for the system to work; it also ensures that there is no seigniorage in the money generation process.

Graziani (1990) stated the conditions for money's existence:

> In order for money to exist, three basic conditions must be met:
>
> a. money must be a token currency (otherwise it would give rise to barter and not to monetary exchanges)
> b. money has to be accepted as a means of final settlement of the transaction (otherwise it would be credit and not money)
> c. money must not grant privileges of seigniorage to any agent making a payment.
>
> The only way to satisfy those three conditions is to have payments made by means of *promises of a third agent*, the typical third agent being nowadays a bank. (p. 11)

Graziani concluded: "any monetary payment must therefore be a triangular transaction involving at least three agents: the payer, the payee, and the bank" (p. 11).

In the process of money creation, the three notions of credit should not be confused: (1) the notion of money as purchasing power, (2) the notion of a bank deposit as credit of the depositor with the bank, and (3) the creation of a bank deposit after taking out a loan. We call "credit money" the money that is identified with bank deposits. Today bank deposits are the debts of the bank with the owners of the bank deposits. A bank owes the holder of a deposit the equivalent in base money, banknotes, or coins. As mentioned, the hierarchy of money has coins and banknotes at its highest level, bank deposits at a lower level. Should banknotes ever be suppressed and only electronic payments accepted, a bank deposit would become a simple statement of purchasing power; the hierarchy of money would shrink to only one level.

The process of the creation of bank deposits is a key element of any modern theory

of money. As remarked in Turner (2013), Wicksell (1898) was the first clear statement of the fact that credit expands the purchasing power of economic agents. This notion is now accepted by most economists, with the exception of those few who continue to argue that money should be based on gold.

There are two opposing approaches to the generation of money as credit, called the exogenous (or verticalist) approach and the endogenous (or horizontalist) approach. The verticalist approach is the classical textbook approach to money generation, espoused by the U.S. Federal Reserve and explained in its booklet *Modern Money Mechanics* (Federal Reserve Bank of Chicago 1992). Let's start with the verticalist approach, which is based on the notion of exogenous money:

> *The verticalists maintain that there is an exogenous source of base money which is then multiplied by the system of commercial banks as credit. The multiplier factor is under the control of the central bank and therefore, in the verticalist approach, central banks control the quantity of money and let interest rates fluctuate freely.*

The Federal Reserve Bank of Chicago's booklet *Modern Money Mechanics* describes the money generation process according to the multiplier. Though not a realistic description of how money is generated by the current banking system, the multiplier helps clarify a number of misconceptions in the press and in textbooks. Let's therefore look at the multiplier.

5.3.1 The verticalist theory of money: banks as intermediaries

The verticalist approach is based on the notion that banks act as intermediaries, granting loans as a fraction of the deposits they receive. A comprehensive exposition of the different theories of banking can be found in Werner (2016) where empirical data on banking practices are also presented. According to the verticalist theory, a bank needs to receive a deposit before it can make a loan. This is the original notion of banks as it developed in Ancient Roman times and successively in the late Middle Ages in Italy and in Germany and which is still widely adhered to today.

The role of banks as intermediaries is an evolution of the role of medieval goldsmiths who offered safekeeping services for the precious metal. Persons holding gold, including gold coins, took the gold to goldsmiths and received in return a certificate of deposit. The depositor remained the owner of the gold, goldsmiths were the safe-keepers.

Box 5.1 The Multiplier

The notion of the multiplier is typical of the verticalist approach to money. Ignoring coins, there are two types of money: banknotes and bank deposits. The banking system is two-tiered, formed by the central bank and by commercial banks. The central bank creates banknotes, the commercial banks create deposits.

Commercial banks work under the fractional reserve system, that is, they keep a fraction of deposits as reserves. Suppose a bank receives a deposit of banknotes. The bank must keep a fraction r of the banknotes as reserves, and can lend the remaining fraction $1-r$. These banknotes end up as deposits in another bank and the process can be repeated ad infinitum.

At the end of the process, commercial banks will have created deposits worth $1/r$ and the banknotes will be kept as reserves. By fixing r the central bank can control the amount of money in circulation. Note that only banknotes can be lent and multiplied.

Banknotes as base money are critical to the multiplier approach. Banks receive deposits in the form of banknotes. If a client of Bank A deposits a check from Bank B, Bank B must transfer banknotes to Bank A. The transfer is not physical but is done through the bank's reserves in the central bank. When a bank makes a loan, it credits a corresponding amount in banknotes to the client who has been accorded the loan. Deposits can be transferred between banks but the total amount of deposits does not change through transfers.

As clients rarely asked to withdraw their gold, goldsmiths and early bankers soon understood that they could safely lend a fraction of the gold they had received in deposit. In order to do so, the legal relationship with the client had to change. After receiving a deposit, the goldsmith (as banks were later to do) released not a certificate of deposit but an IOU by which the goldsmith or the bank became the debtor of the same amount of gold they had received and, simultaneously, took possession of the gold.

Following the introduction of banknotes as legal tender in Europe in the late 17th century, banks accepted not only gold coins but, primarily, banknotes as deposits. When a client deposits banknotes in a bank, the bank becomes the owner of the banknotes and issues an IOU to the client. The client therefore becomes a creditor to the bank and the bank is free to dispose of the banknotes it has received, albeit with a debt to the depositor.

Let's summarize the essence of banking when banks serve as intermediaries:

- The holder takes coins or banknotes to the bank;

- The bank writes an IOU and takes possession of gold or banknotes;
- The bank lends the coins or banknotes now in its possession.

The notion of banks as intermediaries who lend the deposits they receive is a classical textbook description of the banking business. Though it does not correspond to the current practice of the banking systems, it remains the pillar of the notion of the multiplier, which we will now discuss.

5.3.2 The multiplier

In its simplest form, the theory of the multiplier states that the banking system receives exogenous deposits in the form of base money, typically banknotes or coins, and then, by making loans, creates an amount of new deposits whose total value is a multiple of the initial deposit. In order to offer a precise understanding of the functioning of the money generation process with the multiplier, let's initially suppose that there are many commercial banks but there is no central bank (which, remember, was the case in the U.S until 1913). Let's assume that coins and banknotes are issued directly by the government. This assumption (which will be discussed later) is made here for didactical reasons: we need to assume a source of exogenous money.

There are two types of money, that is, two means of payment in the system:

- base money formed by banknotes and coins;
- bank deposits convertible into coins or banknotes.

For simplicity, assume there are only banknotes, no coins, whose total value is in any case small and which, from the point of view of the multiplier, behave as banknotes. Suppose a bank receives a deposit in the form of banknotes. The bank becomes the owner of the banknotes and the depositor becomes a creditor of the bank, with the right to ask for his/her money back in the form of banknotes. In general, there are various agreements between banks and their clients that specify the conditions under which a client can ask for his or her deposit back. In particular, banks might assume the obligation to return the money to clients on demand or after a specified period of time. In general, banks pay interests on deposits in function of the conditions of the deposit. Demand deposits (checking accounts) generally pay no interest.

The idea of *fractional reserves banking* is that banks receive deposits of banknotes and then keep a *fraction* of the deposits as reserves but can lend the

remaining fraction. The rationale of fractional reserves is that, in general, clients will not withdraw all their money at the same time. In the classical tradition, reserves are used to satisfy clients' requests for withdrawing banknotes, the amount of reserves being fixed by the government or by the central bank.

How does the multiplier work? Let's make an example. Suppose that, to simplify calculations, banks must keep 10% of deposits as reserves. Suppose further that a bank – Bank A – receives a deposit of $1,000 in banknotes. The bank creates a deposit of $1,000, that is, it writes a note to the client stating that it has a debt of $1,000 with the depositor, and acquires ownership of the banknotes.

The $1,000 in banknotes are now in the possession of Bank A, which keeps 10% of the deposit, that is $100, as reserves in its vaults and can extend up to $900 in loans. Suppose Bank A lends all the $900. This money, in the form of banknotes, ends up as deposits in the banking system, either in the same Bank A or in another bank. (Indeed it is unlikely that a person would ask for, or a bank would extend, a loan without a project for its utilization.) Suppose the $900 were all deposited in the account of a client of Bank B.

Now Bank B can do the same as Bank A. Bank B will keep as reserves 0.1 × 90(= 90 dollars and will lend 0.9' 900 = 0.9' 0.9' 1000 = 0.9^{2}' 1000 = 810 dollars. This process can continue indefinitely. Assuming the number of banks and the number of clients is infinite, deposits and reserves form two infinite series (see Appendix 5.A for mathematical details):

$$\text{Deposits: } \sum_{i=0}^{\infty} 1000 \times 0.9^{i} = 1000 \sum_{i=0}^{\infty} 0.9^{i} = 1000 \times \frac{1}{1-0.9} = 10{,}000$$

$$\text{Reserves: } \sum_{i=0}^{\infty} 1000 \times 0.9^{i} \times 0.1 = 1000 \sum_{i=0}^{\infty} 0.9^{i} \times 0.1 = 1000 \times \frac{0.1}{1-0.9} = 1000$$

According to the multiplier theory then, with an original deposit of $1,000 in banknotes, the banking system was able to create a total of $10,000 of new deposits. These deposits are formed by the initial deposit of $1,000 plus a total of $9,000 new deposits related to an equal amount of loans. The original $1,000 deposited in Bank A are now distributed as reserves throughout the banking system, creating, as Said, $9,000 of new money as deposits. The banking system has therefore multiplied the initial $1,000 by a lactor of $10 : \frac{1}{1-0.9} = \frac{1}{0.1} = 10,$, that is, by the reciprocal of the fraction of reserves.

A few comments are in order. The newly created $10,000 corresponds exactly to the same amount of loans that have been created as a large number of clients have taken loans. In our previous reasoning the number of loans that could be created is

infinite; in practice, though, only a finite number of loans can be made. $10,000 of newly created money is an upper bound, the real multiplier will be inferior to 10. If there are n clients, the multiplier is $\frac{1-(1-r)^n}{r}$. In the idealized case of an infinite number ot banks, the multiplier is $\frac{1}{1-r}$. See Appendix 5.A tor a proof of the calculations.

The original $1,000 in banknotes are now in the vaults of banks as reserves. Should clients want to withdraw banknotes, banks supply the banknotes from reserves, debiting an equivalent amount from the holders' accounts.

Money thus created as bank deposits can circulate as means of payment in the form of checks. The holder of a check can deposit it in a bank and be credited by the bank. For example, a client of Bank B with a deposit of, say, $900, can write a check of $100, then give it to someone who might be, or become, a client of Bank A. Bank A will credit $100 to its client and then will cash the check with Bank B. Bank B has to supply Bank A with $100 in banknotes from its reserves. This means that banknotes will be physically transferred from Bank B to Bank A. As a result, the total amount of deposits will still be equal to the amount of loans, but there will be clients with deposits and no loans and clients with loans and no deposits or deposits of value less than the loan. The entire process is based on the physical transport of banknotes, as money in the form of banknotes must be physically transported from one bank and deposited in another.

Note that the total amount of money that was created is $9,000. The initial $1,000 in banknotes form the first deposit and are now held as reserves. If a client needs banknotes, he/she will convert some of his/her deposit into banknotes. The sum of deposits will be reduced, reserves will be reduced, and some banknotes will now circulate.

Of course, due to the multiplier, if too many clients want to withdraw banknotes (a situation known as a bank run), banks cannot satisfy them and may eventually be forced into bankruptcy. In addition, if too many clients want to make payments thus transferring their funds to other banks, it might be impossible to satisfy them. A reserve of banknotes is needed both for satisfying direct withdrawals and for making transfers.

Let's now introduce a central bank. Each bank has an account with the central bank. The central bank issues banknotes. When a commercial bank receives a deposit in the form of banknotes, it writes an IOU to the client creating the corresponding deposit, takes possession of the banknotes, and deposits them with the central bank. The central bank, in turn, accepts the banknotes as deposits by commercial banks and issues IOUs to the depositing banks acknowledging the deposits. Deposits with the

central bank plus cash in the vaults are called reserves.

There are several differences here compared to the setting without central banks. First, central banks issue the banknotes that serve as exogenous money. Assume that there is some mechanism that allocates banknotes to the general public (we will here ignore how banknotes issued by the central bank reach the public and become deposits). Now a central bank has issued banknotes for a total value M and this money has become deposits with a total value of $M/(1$ — r). These deposits are the sum of the initial deposit ot M plus $M\left(\frac{1}{1-r}-1\right)=M\frac{r}{1-r}$ deposits associated with loans. In principle the central bank should be willing to print the equivalent of $M\frac{r}{1-r}$ in new banknotes should clients want to convert their deposits into banknotes. In other words, central banks are lenders of last resort: they stand ready to give banks in difficulty as much cash as needed. This facility should prevent banks collapsing from bank runs. Of course, there are physical limits to the amount of cash central banks can provide in a short period of time.

As clients use their accounts to settle payments, commercial banks receive deposits from and lose deposits to other banks. At the end of each day, banks settle their balance of payments among themselves. With a central bank, the settlements between banks within the system are implemented through debiting/crediting reserve accounts as opposed to actual physical transfers. These reserve accounts with the central bank are the equivalent of cash in the vaults of the central bank. They are commercial banks' risk-free deposits, accessible only to the banks themselves.

Commercial banks' lending is limited by the amount of reserves they have with the central bank. Statutory reserves are the reserves that each bank must keep with the central bank in function of the amount of deposits. At each moment, a bank will have an amount of reserves blocked as statutory reserves and an amount of free reserves. Suppose, as before, that statutory reserves are 10%, that is $r = 0.1$. A bank, at any moment, can lend a maximum of $(1 - r) = 0.9$ times the amount of free reserves they have.

In creating bank deposits, most banknotes are taken out of circulation and become reserves or stay in the vaults of banks. The multiplier is thus a process that replaces banknotes with deposits. Central banks print and distribute banknotes that, with time, end up in commercial banks' vaults or go back to the central bank as the reserves of commercial banks. The banks' clients might, of course, withdraw banknotes but the bulk of payments are made through bank transfers.

A critical point is that there must be an original distribution of banknotes to the public to allow them to make deposits. How does the public get the banknotes? Most textbooks take it for granted that the public has banknotes to deposit. If central banks

can finance the state, the public might get banknotes when performing services for the government. However, in many states, including the U.S., the U.K., and the E.U. central banks are not allowed to lend money to the state directly. If central banks cannot lend money directly to the state, the public (individuals or firms) has no way to obtain banknotes in payment of services. Central banks might use banknotes to purchase assets owned by commercial banks. These assets become reserves of the commercial banks who can then make loans to the public who thereby get banknotes. However, the purchase of large amounts of assets by the central bank is an exceptional measure which cannot be considered the standard way of injecting banknotes into the economy.

The vertical approach to money creation is subject to this difficulty: It needs a source of seed exogenous base money to create deposits but it is not clear how individuals and firms can get this seed money if the central bank cannot finance the state.

By controlling the amount of reserves and the amount of banknotes issued, central banks can control the amount of money in circulation. This is the key to the verticalist approach: central banks control the stock of money and let interest rates fluctuate.

5.4 The creation of bank deposits and endogenous money generation

Though the notion of the multiplier discussed earlier is routinely taught in textbooks on banking and financial economics, it does not correspond to the reality of today's banking systems. The alternative explanation of today's banking system and the creation of money – the horizontalist approach – argues that money is created horizontally within the banking system, with the contemporaneous creation of loans and deposits. That is to say, commercial banks can create deposits without actually receiving any exogenous money.

The American economist Basil Moore (1988) played an important role, first in developing the theory of horizontalism and then in carrying out an empirical study which documented that today's banking system does not operate according to the theory of the multiplier. The horizontalist theory of money as being created horizontally within the banking system, thus endogenously, can also be referred to as the endogenous money generation process.

Moore's ideas are gaining acceptance. Explaining the modern process of money creation, McLeay et al. (2014a, 2014b) of the Bank of England's Monetary Analysis Directorate identified as the critical difference between the multiplier theory and endogenous money generation theory their respective explanations of the role of reserves and the role of banks as intermediaries. They wrote that according to the latter theory (1) banks are no longer intermediaries that need deposits to make loans and (2) reserves do not constrain the process of making new loans.

Ulrich Bindseil and Philipp J. König (2013) summarize the debate thus:

> The "verticalist" view states that the money supply function is exogenous, is independent from money demand, and can, at least to a reasonable extent, be controlled by the central bank. The verticalist paradigm may apply in a world of commodity or pure fiat money. But, as Moore argued in 1988, it does not provide a correct description of a credit economy. Rather, in such a world, a horizontalist view must be adopted. The supply of credit money is endogenous, is demand-determined, and only its price can be controlled by the central bank, not its quantity. This was Moore's message in 1988 and it was a brave one.
>
> *(p. 384)*

The critical innovation proposed by Moore – the reason why Bindseil and König call his message "brave" – is the notion of *endogenous* money generation. Horizontalism rubs against a pillar of traditional banking theory: the ability of central banks to control directly the amount of money in the economy. With endogenous money creation, the amount of money in circulation is determined by the market, with central banks controlling directly only the price of money, that is, interest rates.

It would be a mistake to suggest that the process that we are about to explain is the "true" process of money generation. Other systems are logically possible, including the multiplier, and some of them might be implemented in the future. Money is a social technology, linked to societal institutions. Endogenous money generation is the way money is currently generated in advanced economies.

5.4.1 The horizontalist theory of money: endogenous money generation

Let's now explain the process of endogenous money generation. Both the verticalist approach of the money multiplier and the horizontalist approach of endogenous money generation assume the same two-tier banking system formed by commercial banks and a central bank, the latter being the bank of banks. The central bank has an account open for each commercial bank whose deposits with the central bank are called reserves. Reserves are a special form of money exchanged between commercial banks or between commercial banks and the central bank.

In aggregate, commercial banks cannot affect the total amount of reserves; only the central bank can do so, increasing the global amount of reserves either by buying assets from commercial banks or by lending reserves to commercial banks. If the central bank buys assets from a commercial bank, it credits the amount of the sale to the reserve account of the commercial bank. The new reserves are created *ex nihilo*, the central bank simply changing the value of the reserve account of the bank from which it bought the asset. A central bank can reduce reserves performing the opposite operation, that is to say, selling assets to commercial banks or through the repayment of a reserve loan.

New money is created when a client requests a loan. If the bank judges the client creditworthy, it simply creates contemporaneously the loan, which becomes an asset of the bank, and credits the corresponding deposit in the client's account, which then becomes a liability of the bank. With computers, money creation consists simply in changing the value of the client's account.

In both approaches – the multiplier and endogenous money creation approach – a commercial bank creates a deposit *ex nihilo*. The key difference is that in the multiplier approach, a bank can make a loan only if it has sufficient reserves, while in the horizontalist approach the bank can make loans and create deposits without being constrained either by reserves or by funds deposited by clients.

Furthermore, in the horizontalist approach the creation of loans and deposits are processes that do not require any base currency such as banknotes. The condition that deposits be redeemable into banknotes, on demand or with some delay, is not

intrinsic to the process of money creation: it would be possible to create loans and deposits even if banknotes did not exist. However, a mechanism to fix the value of money would be needed to ensure coherence between all bank deposits.

Compare this with the verticalist approach in which base money – the hard currency formed by coins and banknotes – is the essential exogenous money. Base money is produced by the central bank, reaches the public, is deposited in the accounts of commercial banks that act as intermediaries, and goes back to the central bank as reserves of the commercial banks. Only a fraction of the money that is used by the public for payments does not return to the central bank as reserves. Clearly then, the role of base money is an essential difference between the verticalist and the horizontalist approaches. In the former, base money is essential; in the latter base money is not necessary, and in particular not necessary for the creation of bank deposits.

Conceptually there is an intermediate possibility between the verticalist and horizontalist approaches wherein commercial banks are not intermediaries, and produce money endogenously (making hard currency unnecessary) but with reserve requirements reinforced. This would be a system different from the vertical system of textbooks that is based on exogenous money. Similarly, the constraint of bank reserves could be reinforced in the horizontalist approach, with central banks imposing a ratio between loans and reserves even if reserves are no longer linked to banknotes but created as dematerialized deposits by the central bank.

Nevertheless, the notion that banks are constrained by reserves in making loans and in transferring funds does not correspond to the reality of banking today. Indeed, some countries such as Canada have zero reserve requirements. More importantly, central banks always accommodate the demand for reserves: commercial banks make loans of a magnitude dictated by business considerations and then ask their central bank to lend them the requisite reserves if they do not have enough.

Commercial banks can generate an almost arbitrary amount of money. Today the constraints to money generation are business considerations on the profitability of the loans and risk regulations that constrain the aggregate size of loans in function of the capital of banks, not the amount of reserves. The recognition that reserves come after making loans is relatively recent in the literature even if the banking system has operated in this way for decades. The notion of the multiplier, which ties the possibility of making loans to the amount of available reserves, becomes obsolete in a financial system where banks can create any amount of deposits, make the corresponding loans, and turn to the central bank to lend reserves if required either by regulations or by the need to make transfers.

5.4.2 The nature and function of reserves

In the horizontalist approach reserves are deposits – not necessarily of banknotes – of commercial banks with the central bank. Reserves of a bank are increased essentially in the following ways:

- The bank deposits banknotes with the central bank;
- The central bank buys assets from the commercial bank;
- The central bank lends reserves to the commercial bank;
- A commercial bank receives reserves from another commercial bank via a transfer of funds;
- A commercial bank borrows reserves from another commercial bank.

When a bank makes a loan and creates a deposit, its reserves are not changed. Suppose that client X has a deposit of $100,000 with Bank A. Suppose now that client X wants to transfer $50,000 to client Y of Bank B to make a payment. To make the transfer, the account of client X is reduced by $50,000 and the reserves of Bank A are also reduced by the same amount. The reserves of Bank B are increased by $50,000 and the account of client Y is credited $50,000.

Bank reserves are used primarily to make transfers between banks, with each bank receiving money from and transferring money to other banks. Because the balance of transfers in and out is generally small with respect to the total volume of transfers, each bank needs to keep only a small fraction of the entire volume of transfers as reserves with the central bank to be able to satisfy all requests of transfers of funds to other banks. If a bank does not have sufficient reserves with the central bank, transfers cannot be made. In this situation a bank has to borrow reserves from the central bank at a cost or sell assets.

Simon Gray (2011) provides a detailed analysis of the reserve requirements of banks globally. While major economies tend to have similar reserve requirements, there is a wide variety of reserve requirements among central banks worldwide. Noting that the role of reserve requirements has evolved significantly over time, Gray concludes, "The overlay of changing purposes and practices has the result that it is not always fully clear what the current purpose of reserve requirements is, and this necessarily complicates thinking about how a reserve regime should be structured" (p. 1).

Can central banks control the money generation process? In modern financial systems central banks do not control reserves and often, as mentioned in the case of Canada, do not even impose reserve requirements. But central banks do control

interest rates. As reserves in aggregate are money owned by commercial banks and deposited with the central bank, interest rates on reserves are one of the determinants of the profitability of making loans. To increase/decrease the interest rates on reserves, the central bank engages in open-market operations, buying treasury bonds from banks or selling treasury bonds to banks. In buying treasury bonds, the central bank creates reserves for the banks, as banks sell the bonds and are credited the relative amount. By selling bonds it does the opposite. Actually central banks engage in more complicated operations involving repo agreements but the result is the same as buying or selling securities.

5.5 Nonconventional ways central banks can create money

Thus far we have discussed four basic methods of creating money: government minting of coins or printing banknotes, commercial banks issuing loans, and central banks creating bank reserves. But central banks have no direct constraints in the creation of money; indeed they have additional, albeit nonconventional, ways of creating money. One such nonconventional way is through the purchase of assets. If central banks purchase assets from commercial banks they credit the reserve account of the bank from which they bought the asset with a corresponding amount of reserves. (The process through which central banks buy or sell assets is called open-market operations.) In theory, increasing reserves leads to lower interest rates thereby encouraging firms to borrow for investment purposes and banks to make loans.

However, after the 2008 financial crisis firms were not quick to increase their borrowing. As described in Turner (2013) new money is presently generated more from the borrowing of the general public than from the borrowing of firms (see Chapter 8 for details). In the U.S. and the U.K., firms are sitting on piles of near-money (see Chapter 1 for Grantham's explanation as to how S&P 500 firms were able to raise margins and profits and accumulate cash in the post-1995 period) and are now, in aggregate, net lenders through financial markets; when they borrow from banks, it is often to buyback stocks rather than invest. Birinyi Associates estimate that U.S. listed firms spent about $6.1 trillion buying back their own shares during the 11-year period 2005–2016.

As economies failed to pick up following the 2008 financial crisis, central banks adopted the nonconventional operations called Quantitative Easing (QE). QE consists in central banks buying assets (which might include commercial bonds and stocks) from non-banks such as pension funds or insurance firms; it is a direct injection of money into the economy. To see how it works, suppose a central bank decides to buy $10 million worth of bonds from non-bank A. Because central banks do not have relationships with non-banks, the central bank will instruct a corresponding Bank B to buy the bonds from non-bank A, crediting Bank B's account with $10 million. Bank B will in turn sell the bonds to the central bank which will credit $10 million to the reserve account of Bank B. The central bank creates $10 million of reserve *ex nihilo*. The final result is that $10 million of new cash has been created and deposited in the account of pension fund A. Bank B has a new deposit and the corresponding amount of reserves. The impact of QE on reserves is

illustrated in Exhibits 7.5 and 7.6 that show the increase in bank reserves caused by QE.

Other nonconventional ways of creating money *ex nihilo* have been proposed in the past, for example helicopter money, without ever having been implemented. Helicopter money is nonconventional in that money is created by the central banks and distributed uniformly to the public without any loan.

In summary, the endogenous money creation theory holds that money is created by commercial banks without the intervention of the central bank. The central bank intervenes only to provide reserves so that regulatory standards are met and it is possible to make transfers of funds between banks. In this view, banknotes play no role in the money generation process. Banknotes are supplied to individuals on demand but do not influence the total money generation process. An endogenous money generation process can start from zero with banks making loans and creating deposits. This is a fundamental difference with the verticalist theory which holds that the money generation process depends on exogenous money supplied by central banks and multiplied (but not created) by commercial banks. The main difference is given by the handling of reserves and by the role of currency.

5.6 Other ways to create money

In addition to the creation of money or near-money that we have discussed thus far, there are additional ways to create purchasing power, including private credit and the issuing of digital currencies. The purchasing power thus created does not always qualify as money.

5.6.1 Private credit

We have discussed the role of commercial banks in creating money by issuing loans. But commercial banks are not the only ones to accord credit or issue loans. There are two other commonly used ways to accord credit: commercial payment terms and peer-to-peer lending.

In according payment terms to their business partners, firms extend credit to clients, thereby according them increased purchasing power. Consider, for example, clothing stores. Typically shops order goods well in advance of the season when products can be sold and, at the beginning of the season, keep inventory. Modern techniques of flexible production allow just-in-time delivery of products on short notice but, still, any clothing store must keep a minimum of inventory. It makes a difference if the supplier insists on being paid on delivery, within say 60 days, or only in the case of a sale, accepting to take back unsold inventory. By increasing the purchasing power of clients, commercial payment terms are loans and fulfill one of the key functions of money. However, firms that give favorable payment terms are using their own money or reducing their own credit and, ultimately, their own purchasing power. The sum of credit/ debt is zero.

It is common business practice that firms discount their credits with clients. That is, they take the IOUs to their bank and the bank credits the corresponding amount to their accounts. In this case, money is created as the bank extends credit to the firm using the IOUs as collateral. It is therefore clear that the relationship between supplier and client does not create money; rather money is created by the bank when it discounts the IOUs.

Similar considerations apply to private lending. The practice of private lending is very old, with social connotations that vary from generous help to the needy to sheer exploitation through usury. The novelty of peer-to-peer (P2P) or marketplace lending is the use of modern market tools such as the Internet to match borrowers and lenders in a systematic way. Marketplace lending is challenging the role of commercial banks in that individuals or businesses can borrow and lend directly among themselves

online, without commercial banks playing the role of intermediary. Actually, according to a Morgan Stanley Research paper (2015) the fastest-growing platforms are not P2P but institutional investors partnering with technology platforms to cherry-pick borrowers.

An advantage to the lender is that the loan might generate interest in excess of that earned by traditional means such as savings bonds or certificates of deposit (CDs). Marketplace lending might thus offer better storage of value, but it comes with greater risk. Nevertheless, marketplace lending does not create money and does not increase purchasing power in aggregate: it simply transfers money. Here we see the fundamental function of banks – as opposed to private individuals and non-banks – is their ability to create money *ex nihilo*.

But what about other players – some very nontraditional – that are proposing digital cash and digital currencies for use in financial transactions? Do they create money? Recall that near-money assets such as money-market funds are not directly used for payments. Let's begin our discussion with digital cash.

5.6.2 Creating digital cash

There are different forms of electronic cash: credit cards, digital wallets, and payment systems. From the point of view of the theory of money their functions are different. Credit cards are similar to bank credit but (1) the credit is typically supplied by non-banking organizations such as Visa or MasterCard and (2) the credit is extended on demand via an electronic system of reading and checking the cards. Whether credit cards effectively add to the total of credit and increase the supply of money depends on how the circuit is organized. An individual purchasing something with a credit card is using purchasing power provided by the card company. The person acquires some goods and has a debt with the credit card company. If the credit card company immediately pays the supplier, there is an increase in the total amount of money in circulation; if the contrary is true, there is no increase in the money supply.

Digital wallets, also known as e-wallets, are software applications that store personal information of the holder (they are effectively information systems); this information is used to facilitate transactions using payment systems or credit cards. The Google Wallet is an example. Digital wallets have no impact on the process of money generation. Payment systems, on the other hand, are systems to transfer payments without the storing of personal information. Like digital wallets, they have no impact on money generation.

5.6.3 Creating private digital currencies

As discussed in Chapter 3, digital (or crypto) currencies such as bitcoins are not coins at all: they are electronic certificates "bought" over the Internet from "miners" or exchanged for other digital currencies, for example, over digital currency exchanges. (See Chapter 3 for a discussion on the underlying technologies including distributed ledgers and blockchain technology.) New bitcoins are created as remuneration for miners who execute and verify transactions, thereby providing the record-keeping service. Once created, bitcoins can be used to purchase goods and services. End-November 2017 it was reported that there were almost 17 million bitcoins (certificates) in circulation, estimated to be worth about $138 billion.

"Miners" are non-salaried individuals or firms who use complex mathematics and their own powerful computers (upwards of tens of thousands) as they compete to record and verify the estimated 225,000 bitcoin transactions daily. They operate independently and autonomously (though typically in pools) in a decentralized global network whose computing capacity, in early 2016, was estimated to be 43,000 times more powerful than the world's top 500 supercomputers combined. Miners from China are estimated to provide more than two thirds of the bitcoin network's computing power.

Do private digital currencies – electronic commodities that have no intrinsic value – add to the existing pool of money? CoinMarketCap estimated that the total value of digital currencies issued as of August 10, 2017 was about $125 billion and is $138 billion as of end-November 2017. But do digital currencies qualify as money? In Chapter 3 we suggested that digital currencies do not qualify as money: they are not very liquid, do not convert at par, and their acceptability as a means of payment is presently limited. In addition, digital currencies as they are created today violate the fundamental principle of modern money creation that prescribes that no private entity can issue money for their own benefit. Bitcoins are issued by miners who are paid in the currency they created. Can one trust a monetary system that benefits private issuers? As it is, central banks, whose role is to maintain trust in the monetary system, are now beginning to consider the supervision of these currencies.

5.6.4 Creating state-issued digital currencies

Some central banks are actively researching the possibility of issuing their own digital currency based on blockchain technology. Among them are the Bank of England, with the Ripple interledger experiment conducted together with the

California-based blockchain specialist Ripple. Another is the Bank of Canada (see Chapter 3 for how Wilkins and Gaetz [2017] evaluated the Bank of Canada's experimental wholesale interbank payment system using distributed ledger technology). Ben Broadbent (2016), deputy governor for Monetary Policy at the Bank of England, remarked that the introduction of a central bank digital currency involves more than a narrow, technical judgement about the efficiency of the payments system. Adopting distributed ledger techniques for the generation of digital currencies would, he said, be a major change, opening the possibility that the public hold accounts directly with the Bank. A particular concern at the Bank of England: the extent to which a central bank digital currency would compete with the main form of money in the economy and commercial bank deposits. In any case, should central banks eventually issue their own digital currency, the creation of the currency will surely be realized using criteria different from those of today's private issuers.

Appendix 5.A The equations of the multiplier

Having discussed the multiplier, let's now explain the mathematics behind it. If the required reserve ratio is $r < 1$, assuming an infinite number of banks, an initial deposit C is multiplied into an infinite number of deposits with a total value of. $\frac{C}{r}$. For example, if the reserve requirement is r = 0.1, an initial deposit of $1,000 is multiplied into a total of $\$\frac{1000}{0.1} = \$10{,}000$

In fact, the initial deposit of C in a bank A_1 allows the bank A_1 to loan $C(1-r)$. This loan will likely be deposited in a bank A_2 which can make in turn a loan of $C(1-r)(1-r) = C(1-r)^2$. In general, the bank A_n will receive a deposit of $C(1-r)^{n-1}$ and will make a loan of $C(1-r)^n$.

Consider the first n banks. They will be able to create the following deposits:

$$\begin{aligned} C_r^1 &= C(1-r)^0 = C \\ C_r^2 &= C(1-r)^0 + C(1-r)^1 \\ &\vdots \\ C_r^n &= C(1-r)^0 + \cdots + C(1-r)^{n-1} \end{aligned}$$

To compute the value C_r^n we can establish two simple recursive relationships:

$$C_r^{n+1} = C_r^n + C(1-r)^n$$

$$C_r^n(1-r) + C(1-r)^0 = C(1-r)^0 + (1-r)\left[C(1-r)^0 + \cdots + C(1-r)^{n-1}\right] = C_r^{n+1}$$

Combining the two we obtain:

$$C_r^n + C(1-r)^n = C_r^n(1-r) + C$$

$$C_r^n = C\frac{1-(1-r)^n}{1-(1-r)} = C\frac{1-(1-r)^n}{r}$$

Therefore, the sum of the first n deposits is:

$$C\frac{1-(1-r)^n}{r}$$

If we consider an infinite number of banks, given that $(1-r) < 1$ we can write:

$$C_r^\infty = \lim_{n\to\infty} C\frac{1-(1-r)^n}{r} = C\frac{1}{r}$$

Notes

1 During the period of the Viking raids (*c*.800–1000) slaves were Ireland's most valuable exportable commodity. A similar situation existed in eastern Europe (see Wickham 2010 [2009], p. 496 in Penguin's 2010 edition).

2 Actually, coins have also been made of basic minerals that are either not readily available or difficult to extract from the earth, such as salt. Cakes of salt were used as money in central Africa. Returning from Cathay in 1295, Marco Polo is reported to have entertained the Doge Pietro Gardenigo with tales of sa coins of high value bearing the seal of the great Mongol sovereign Genghis Khan.

3 Source: http://mentalfloss.com/article/12715/why-did-us-abandon-gold-standard.

4 Source: Museum of the National Bank of Belgium: www.nbbmuseum.be/en/2007/09/chinese-invention.htm.

5 For the full story on free banking in the U.S in the 19th century, see Galbraith (1975), Chapter 8, "The Grea Compromise."

6 The second largest owner of gold is Germany, with 3,387 tons of gold, a GDP of 3,467 trillion USD (2016) and M2 of over 3,108 trillion USD in May 2017.

References

Bindseil, Ulrich and Philipp J. König. 2013 . "Basil J. Moore's Horizontalists and Verticalists: An Appraisal 25 Years Later." *Review of Keynesian Economics*, vol. 1, no. 4 (Winter):383–390.

Broadbent, Ben. 2016. "Central Banks and Digital Currencies." Speech given at the London School of Economics. (March 2): www.bankofengland.co.uk/publications/Pages/speeches/2016/886.aspx.

Federal Reserve Bank of Chicago. 1992. *Modern Money Mechanics: A Workbook on Bank Reserves and Deposit Expansion.* Last revision 1992, reprint lulu.com, 2011.

Galbraith, John Kenneth. 1975. *Money: Whence It Came, Where It Went*. New York: HarperCollins.

Goodhart, Charles. 1998. "The Two Concepts of Money: Implications for the Analysis of Optimal Currency Areas." *European Journal of Political Economy*, vol. 14, no. 3:407–432.

Graeber, David. 2011. *Debt: The First 5,000 Years*. New York: First Melville House.

Gray, Simon. 2011. "Central Bank Balances and Reserve Requirements." IMF Working Paper. (February).

Graziani, Augusto. 1990. "The Theory of the Monetary Circuit." *Economies et Societies*, vol. 24, no. 6:7–36.

Ingham, Geoffrey. 2004. *The Nature of Money*. Malden, MA: Polity Press.

McLeay, Michael, Amar Radia, and Ryland Thomas. 2014a. "Money in the Modern Economy: An Introduction." *Quarterly Bulletin* (Q1). Monetary Analysis Directorate, Bank of England.

McLeay, Michael, Amar Radia, and Ryland Thomas. 2014b. "Money Creation in the Modern Economy." *Quarterly Bulletin* (Q1). Monetary Analysis Directorate, Bank of England.

Mehrling, Perry. 2015. "The Inherent Hierarchy of Money." Chapter 21 in *Social Fairness and Economics*, edited by Armon Rezai, Lance Taylor, and Tom Michl. Abingdon: Routledge.

Moore, Basil. 1988. *Horizontalists and Verticalists: The Macroeconomics of Credit Money*. Cambridge, UK: Cambridge University Press.

Morgan Stanley Research. 2015. "Can P2P Lending Reinvent Banking?" Morgan Stanley Blue Papers. (June 17): www.morganstanley.com/ideas/p2p-marketplace-lending.

Turner, Adair. 2013. “Credit, Money and Leverage: What Wicksell, Hayek and Fisher Knew and Modern Macroeconomics Forgot.” Speech made at conference “Towards a Sustainable Financial System,” Stockholm School of Economics. (September 12): https://cdn.evbuc.com/eventlogos/67785745/turner.pdf.

Werner, Richard A. 2016. “A Lost Century in Economics: Three Theories of Banking and the Conclusive Evidence.” *International Review of Financial Analysis,* vol. 46:361–379.

Wickham, Chris. 2010 [2009]. *The Inheritance of Rome: A History of Europe from 400 to 1000*. London: Penguin Books [Allen Lane].

Wicksell, Knut. 1898. First published as *Geldzins and Guiterpreise*. English version 1936, *Interest and Prices*. New York: Macmillan.

Wilkins, Carolyn and Gerry Gaetz. 2017. “Could DLT Underpin an Entire Wholesale Payment System?” (Contributed to) *The Globe and Mail*. (May 25): https://beta.theglobeandmail.com/report-on-business/rob-commentary/could-dlt-underpin-an-entire-wholesale-payment-system/article35106771/

6

How Money Acquires Value and How that Value Changes Over Time

How does money gain acceptance? How does it acquire value? How (and why) does that value change over time? Let’s start with the first question: How does money gain acceptance?

6.1 How money gains acceptance

In trying to address the question as to how money gains acceptance, a distinction is called for between money with an intrinsic value and fiat or credit money. If money has intrinsic value, as in the case of gold or silver coins, its acceptance is related to its intrinsic value strictly matching its face value: should the value of the metal be higher than the coin, coins would be melted and disappear; should the value of the metal be only a fraction of the nominal value of the coin, then we are back in a situation where trust is a needed ingredient for the coin's acceptance.

Let's discuss the case where money is fiat (or credit) money – the current situation in most developed economies. Money circulates and has value because people believe that (1) it will be accepted by other people and (2) it will not lose value. Money is by nature a social tool: any specific form of money has value if a sufficiently large community accepts it; consider local currencies, discussed in Section 3.2.2. Today, given that in most advanced economies money has no intrinsic value, acceptance of money is based on trust in the institutions that guarantee money: the government and a banking system consisting of a central bank and regulated commercial banks.

Trust is not a given. Consider the 2002 peso crisis in Argentina following that country's Great Depression (1998–2002): a lack of trust saw Argentinians refuse state money in favor of other currencies, in particular the USD. In other cases, such as the Swiss WIR Franc, though the national government does not suffer from a lack of trust, local money is preferred by a close-knit community of enterprises and individuals.

So how is trust in money achieved? Two main theories have been proposed. The first is due to Carl Menger. Born in 1840 in Neu-Sandez in Austrian Galicia, now in Poland, Menger founded the Austrian School of Economics. Menger (1892) advanced the idea that money is a social tool but argues that its acceptance must grow endogenously in the market. Menger's theory of how a form of money gains acceptance is the following. Money evolved from barter. Things have different marketability and therefore barter opportunities are easier to find for some goods than for others. Gradually precious metals emerged as the most marketable of goods and successively started to circulate as coins. According to Menger, money is the product of exchange in the private sector, that is, money is a product of the initiative of market participants and its acceptance therefore grew out of its function in the private sector.

The other main theory of how money is accepted – the State Theory of Money -

argues that money is accepted due to the intervention of the state. According to the State Theory of Money, state-issued money is accepted because the government requires state money in payment for taxes and other sums due and uses state-issued money in payment of goods and services. Because flows in and out of governments represent a large fraction of the economy, the state currency becomes the de facto monetary standard. In Western economies, 2015 government spending as a total of the GDP ranges from a low of 38% for the U.S. to a high of almost 57% for France; other major developed countries fall somewhere in between.[1] Figures from the World Bank for 2015 general government financial consumption expenditures as a percentage of GDP are somewhat lower but the relative ranking of states is similar, with the U.S. government financial consumption expenditures representing roughly 14% of the nation's GDP and France on top with government financial consumption expenditures representing roughly 24% of the nation's GDP; the figures for Germany, Italy, and Japan are in the 20–30% range.[2]

The State Theory of Money, proposed by the German economist Friedrich Knapp (1905 in German, English translation 1924), argues that money has no intrinsic value, but acquires value through the state. This view is known as Char-talism. Knapp wrote the State Theory of Money in 1905, when virtually all major currencies were on the gold standard. His theory was opposed by the Metallists who argued that money is, and must be, something with an intrinsic value.

Still other theories have been proposed. Ludwig Von Mises (1912; 1st English translation 1934) suggested that money gains acceptance if people believe that it can be used in the near future to buy the essentially same things that it can buy today. Von Mises argued the reasonableness of this belief: If we have seen, from the immediate past, that the money has conserved its value, we can reasonably assume that it will retain its value in the near future. (Of course anyone with a different experience might disagree.) Nevertheless, Von Mises observed that this reasoning is not an infinite regress: going back in time, we will find a moment when money had an intrinsic value as gold or silver. Von Mises supported the gold standard and argued that the value of money is ultimately related to something with intrinsic value.

Box 6.1 Is Money Being Privatized?

Central banks provide fiscal policies, control interest rates, ensure the transfer of funds, and issue base money that serves as the unit of account. But as regards the generation of new money, most new money is generated by commercial banks when they create credit/debit pairs. Bank deposits are ultimately a

private currency linked to the state money by well-defined rules.

The emergence of cryptocurrencies and of other complementary currencies can be seen as parallel private currency systems related to the primary currency system by an exchange rate. Today currencies such as bitcoins, Ethereum, or Ripple, to name a few of the more than 840 digital currencies on the market today, are essentially a rare commodity sold at a market price. But there are also private currency systems such as the local Swiss WIR Franc founded by the Basel-based WIR Bank during the Great Depression and still used today by 62,000 Swiss small and medium-sized enterprises (see Chapter 3).

It is impossible to predict if the trend towards the creation of private digital currencies and the development of virtual communities where transactions are made with different private currencies will continue. Clearly the fate of these private currencies is with governments and central banks, which are beginning to take the phenomenon seriously. The People's Bank of China has announced its intention to supervise these currencies and to create a state-issued digital currency. The People's Bank of China is not alone; other central banks including the Bank of Canada and the Bank of England are also exploring the issue (see Chapters 3 and 5). As we saw in Chapter 3, in the past, governments did not look kindly on alternative private currencies. Attempts in Austria and Germany to create local currencies during the Great Depression were declared illegal.

The debate goes on. On one side, in modern economies the government is certainly a major actor, responsible, as we have seen, for a large fraction of the economy. In addition, governments have coercive power to require state-issued currencies in payment of taxes and other sums. But it is also true that the private system has demonstrated the ability to create and win acceptance of its own money, as Menger argued. The WIR Franc, used as money by one fifth of the Swiss economy, is one example; cryptocurrencies may represent another.

More importantly, state-issued money is only the unit of account; the vast majority of money in circulation in developed economies is formed by bank deposits, created and managed by commercial banks. While governments insure deposits up to a given sum, it is unlikely that individuals and firms trust the private banking system only because their bank accounts can be converted into banknotes and their deposits are guaranteed by the government. This might be true for small savers but certainly not for big organizations that transfer huge amounts of funds daily.

So how did these private organizations that are today's commercial banks earn the

public's trust? Recall that this was not always the case. The U.S. Federal Reserve System was created in 1913 largely because the public, especially the farmers in the Midwest, did not trust organizations based in New York. But to win, in turn, acceptance to the creation of a central bank, it was necessary to create a system of regional Federal Reserves.[3]

The trust in private financial organizations has grown since the early 20th century and into the 21st, despite market crashes and, in some cases, subsequent economic crises.[4] Consider derivatives markets formed by banks and exchanges where close to one trillion dollars are traded daily. A century after the formation of the U.S. Federal Reserve, critical financial functions such as the creation and management of pensions or the management of investments are entrusted to private organizations, banks or non-banks as the case may be. Clearly the question of the acceptance of money (and its managers) is less problematic now than it was at the beginning of the 20th century, when people were reluctant to accept pieces of paper not convertible into gold.

Today in most of the developed world we trust – or behave as if we trust – modern financial systems, banks and non-banks alike. Consider that more than $37 trillion in savings worldwide are managed by investment companies. In the U.S. alone, at year-end 2015, according to Investment Company Institute (2016), more than 90 million U.S. retail investors were entrusting the management of over $18 trillion of their savings to U.S.-registered investment companies.

Today's world is one of big organizations. We can pick up a phone and call virtually anywhere, board a plane in one airport and reach virtually any other airport in the world – a considerable feat of logistics. Every time we push the brake pedal of our car, we implicitly entrust our lives to a large engineering and manufacturing organization. At the time Knapp and Menger wrote, the world of ordinary people was much smaller than it is today. Only governments and the military were organized on a large scale. The world of big organizations really started in the second half of the 19th century with the development of railroads, electricity, and the telegraph. The trust that people place in money and the banking system is part of a generalized trust in big organizations.

But organizations, like governments, can lose trust. Consider, as mentioned above, the 2002 peso crisis in Argentina, or investor behavior after the 2007–2008 financial crisis. As reported by Nathaniel Popper (2012), despite U.S. stocks having doubled in price over the three-year period 2009–2012, trading in U.S. stock markets continued to fall – from a peak of 12.1 billion average trades a day in 2008 to 6.5 billion in April 2012. This decline, Popper wrote, "stands in marked contrast to the past economic recoveries, when Americans regained their taste for stock trading within two years of economic shocks in 1987 and 2001." Perhaps the six distinct

shocks in the U.S. markets during the past 30 years or so that David Romer (2013) counted shook (at least until April 2012) Americans' trust in the stock markets (see note 4).

6.2 How money acquires value

"The value or purchasing power of money depends, in the first instance, on demand and supply . . . The supply of money . . . is all the money in *circulation* at the time . . . The demand for money, again, consists of all the goods offered for sale." Thus wrote John Stuart Mill (1852, pp. 12–13; italics in original).

How does money acquire value (purchasing power)? Money is a unit of account. But how is the unit of account established? How is it that a $100 banknote has the purchasing power it has? Observe that prices are relative prices. There is nothing in a price system that intrinsically fixes the absolute value of goods and services. So how does a given currency acquire the purchasing power it has?

Let's make a thought experiment, unrealistic but useful for fixing some ideas. Imagine a community that is working on pure barter, without money. Assume that the community has developed a complete "conversion table" so that the relative value of all available goods is defined. The table will include entries such as "one cow is worth two sheep" and, in our case, also a conversion table of goods and services in terms of metal such as silver or gold.

Now imagine that a government wants to introduce fiat money into this community. The government has first to introduce a currency. Let's call this currency SUN. The government next prints banknotes of various nominal values, say, 10 SUN and 100 SUN. We see here the two aspects of a currency: (1) it is an abstract measure of value and (2) it is implemented as "tokens," that is, banknotes, of a given nominal value. On introducing the SUN the government will say something like: "Starting today every good and service and payment due to the government will have a value measured in units of SUN." In itself this is an abstraction.

The key issue now is how to make these pieces of paper circulate and become the de facto means of exchange. The government must set the purchasing power of its SUN. The decision is fundamentally arbitrary. It might, for example, stipulate that an hour of work is worth one SUN.[5] Or it might decide that a gram of silver is worth one SUN. Of course, in ordering people to work in government projects (such as building a temple) the government must have printed enough SUN banknotes to be able to pay the workers and the corresponding value in goods needed to complete the project. Recall that we assume that relative prices had been previously established in a conversion table before the government introduced banknotes. Workers are paid by the government and then use their money to buy goods and services from others. As approximately every household needs money to pay taxes, every household will be willing to accept SUN money.

This thought experiment illustrates the intrinsic difficulty in "injecting money into the economy," a problem that does not have an obvious solution. How is the rest of the economy going to receive money for their transactions? How do they get enough SUNs for their commercial needs? Clearly more money is needed.

It's time to abandon our thought experiment. From an engineering point of view there are many solutions to injecting money into an economy. In practice, though, two solutions have been adopted. The first consists in the government paying for goods and services with the newly created currency and requiring this in return when collecting taxes and other sums due, as in our thought experiment. The second is lending. A government can lend money to allow exchange. At this point a banking system is needed, as suggested by Graziani (1990) (see Chapter 5). With a banking system and a variable supply of money, a government is no longer able to fix the purchasing power of its money. Over time, the purchasing power of the state-issued currency will adjust to reflect supply and demand. (Note that how the quantity – or supply – of money affects prices is discussed in Chapter 8 where we present the Quantity Theory of Money and discuss its shortcomings, that is, its failure to consider the unequal distribution of money.)

To our knowledge, no government has introduced fiat money by design into a barter economy. As discussed in previous chapters, early commercial transactions were conducted using a combination of barter and credit–debt relationships and later metallic bullion and coins. The abstract notion of value was first linked to bullion and coins, whose value originally derived from the intrinsic value of the metal therein. However, under the pressure of supply and demand, though their purchasing power changed, the unit of account remained. For example, the sestertius remained as a unit of account in the West some 100 years after the collapse of the Western Roman Empire.

It should be clear by now that the purchasing power of a given unit of currency depends on the history of the currency. Modern economies are substantially indebted to their history. How much $1 can buy today depends on the dollar's past. At any moment a government can step in and change the numeraire. For example, in 1960, the French franc was replaced by a new (heavy) franc whose value was 100 times the value of the old franc. This move was part of a bigger financial restructuring of French public finances. In itself, however, it did not have any consequence as both prices and wages were divided by 100. Clearly the value of money is conventional.

As an illustration of the working of supply and demand, consider the recent launch of a major new currency – the euro. The euro was first launched virtually in January 1999 and then three years later, in January 2002, with coins and banknotes. The value of the euro was set as the value of one ECU (European currency unit)[6] which at the

time was trading at 1.1686 ECU/USD. Euro dollar trading started on 31 Decembe 1998 at the euro/dollar exchange rate 1.1686; four days later, on 4 January 1999, the euro closed at about $1.18 but it dropped to parity with the USD by the end of the year. During the U.S. subprime mortgage crisis, the euro reached a peak of $1.6037 in July 2008 but in mid-year 2017 it was trading at about the same value as at its launch more than 18 years earlier, that is, in the range of 1.00 euro = $1.17–$1.18.

Once a parity value has been set, the market intervenes, as we saw above in the case of the euro dollar exchange rate. But to track the evolution of a currency's

value in terms of its purchasing power (the definition of money we gave at the beginning of our discussion) we use inflation indices such as the Consumer Price Index (CPI) for the internal evolution of purchasing power or international indices such as *The Economist*'s Big Mac index for international comparison. Both are clearly inadequate. The Big Mac index (admittedly a lighthearted index) is based on the theory of purchasing power parity,[7] and uses the relationship between prices and GDP per person to evaluate if a currency is under-/over-valued. According to the latest release (July 17, 2017), a Big Mac in the U.S. costs $5.30 and in the euro area $4.47 (€3.91), indicating an implied exchange rate USD/EUR of 0.74 versus the actual exchange rate of 0.88.[8]

Let's now discuss problems associated with using inflation measures to understand changes in the value of money inside a given economy.

6.3 How the value of money changes: the elusive concept of inflation

Money as purchasing power sees its value change over time; that is, what a given unit of money buys today might be different from what it can buy tomorrow. The value of money might go up or down. If it goes down, that is, if prices go up and a unit of the currency buys less, we call the phenomenon inflation.

Inflation is one of the most important concepts in economics; it is a key concept from both the practical and theoretical points of view. It affects people's well-being, determines welfare transfers, and influences monetary policy. From the theoretical point of view, inflation is a critical concept in both macroeconomics and growth theory as it is used to recover "real" quantities from "nominal" quantities. Inflation is also an important concept for any theory of money: should it be possible to measure inflation, inflation would determine how the value of money (i.e. its purchasing power) changes over time.

Superficially, inflation is a concept that seems solid and easy to understand. Consider, for example, the U.S. in the 1950s, when the price of a car ranged from $1,300 for a bottom-end Chevrolet to $14,000 for a Cadillac Eldorado Brougham then the most expensive car in the U.S. market. Today, the price of cars ranges from about $10,000 for a small compact to $500,000 for a Rolls Royce Phantom and upwards of $1 million for supercars such as a Bugatti, Ferrari, Koenigsegg, or Lamborghini. Clearly we can observe that the price level of cars has gone up some 10–15 times in the 65-year period, but we can also observe that the price range in the 1950s was 1 to 10 while the price range today is 1 to 50+. Something big has changed in the economy.

Similar considerations can be made for many other products and services. Anyone who has experienced a period of rapid inflation would say that inflation is a well-defined phenomenon with important consequences for their daily life. If prices go up and wages remain stable, households will find it increasingly difficult to buy the goods and services they were accustomed to buying. If wages go up and prices follow the same trend, people do not experience any change in their purchasing power.

Most people's intuition of inflation is paralleled by the definition of inflation in textbooks, albeit with many caveats. Here are a few examples. Samuelson (1976) defines inflation as follows: "By inflation we mean a time of generally rising prices for goods and production factors – rising prices for bread, cars, haircuts; rising wages, rents etc." But immediately after, Samuelson states that, "Neither in inflation

nor in deflation do prices move all in the same direction or in the same proportions" (p. 27).

For Mishkin (2007), "the average price of goods and services in an economy is called the **aggregate price level** or simply the *price level*" (p. 10; bold and italics in original). In an appendix to Chapter 1, the author tries to make the definition more precise, with reference to three popular deflators: the Gross Domestic Product (GDP) deflator, the Personal Consumption Expenditure (PCE) deflator, and the Consumer Price Index (CPI) deflator, stating that the CPI deflator is the one commonly used.

Other economists including Romer (1996) consider the concept of inflation to be so well known as to not require a precise definition. In his introductory book on inflation, the economist James Trevithick (1977) deems it unnecessary to give a definition of his subject.

On the other hand, Florin Aftalion and Patrice Poncet (1995) define inflation as the continuous growth of the general level of prices.[9] For these economists, inflation is characterized by a *continuous* upward trend in *all* prices; if some prices go up while other prices go down we are not in a state of inflation (p. 55).

Despite the apparently robust intuition, inflation is a difficult-to-define concept. Why? First, prices do not all move together. Some prices go up while others go down, all in different proportions. Second, there is innovation in products and services so that if we consider inflation over a period of time, say a year, at the end of the period consumers might be able to buy products that were not available at the beginning of the period – or alternatively not able to find products that were available at the beginning of the period. Clearly, it is possible to define an infinite number of measures of inflation.

The problem of defining a true index of price levels has been debated since the end of the 19th century without arriving at an agreed-upon index. Different indices gave different results. To solve the problem, Irving Fisher and others devised tests that any index should be able to pass. But the Norwegian economist and founder of econometrics Ragnar Frisch (1930)[10] wrote his famous paper "Necessary and Sufficient Conditions Regarding the Form of an Index Number Which Shall Meet Certain of Fisher's Tests" in which he demonstrated that no index can pass all tests. Actually he demonstrated that no index can simultaneously pass any three key tests.

Alexander Konüs (1939) proposed a theoretically correct (but not feasible) solution based on computing a utility index for each economic configuration of prices and quantities. However, utilities are not empirically observable quantities and, in addition, might change as tastes change. The "economic" approach of Konüs originated a vast literature devoted to defining true indices under different

assumptions of consumer behavior. For a historical perspective on the problem of indexing, see Carlo Milana (2009).

To outline the problem of measuring price levels, consider an economy that produces n goods with prices and quantities $(p_{i,t}, q_{i,t})$, $i = 1, \ldots, n$. The value of the output at time t is $\sum_{i=1}^{n} p_{i,t} q_{i,t}$. If quantities and relative prices do not change, then it would be simple to determine the price level as the common multiplier of all prices. But, in practice, relative prices and quantities do change.

A simple partial solution is to consider only a basket of products on the market for the entire period and evaluate the change in the value of the basket. Three main indices have been proposed. They are:

$$\text{the Laspeyres index, which is written: } P_L = \frac{p_1 q_0}{p_0 q_0};$$

$$\text{the Paasche index, which is written: } P_P = \frac{p_1 q_1}{p_0 q_1};$$

$$\text{and the Fisher index, which is written: } P_F = \sqrt{P_P P_L}.$$

The Laspeyres index keeps the quantities fixed for the period at their *initial* level and changes only prices; the Paasche index keeps quantities fixed for the period at the value they have at the *end* and changes prices; the Fisher index is the square root of the two indices.

Both the Laspeyres and Paasche indices compute an index of price level for a subset of the market formed by products and quantities that remain unchanged. In practical applications, the index is then extended to the entire economy. We can use these indices to compute inflation rates.

Note explicitly that we cannot say that these indices are an approximate estimate of a true inflation index because a true inflation index does not exist. A true inflation index would imply a physical measure of the aggregated output allowing us to compute the change in price of a unit of aggregated output. But no physical aggregation is possible as we cannot physically aggregate vacation packages, cars, telephone and Internet services, and the myriad other products offered in modern economies.

In any modern science there is no place for intuitive a priori concepts. Physical concepts acquire meaning in terms of the entire body of physical laws *and* the observation process. The notion that physical concepts are defined in terms of observations (i.e. physical theories are based on observations) was stated as a

fundamental methodological principle by the American physicist Percy Bridgman, in his 1927 book *The Logic of Modern Physics*. (Bridgman received the 1946 Nobel Prize in Physics for his work on high-pressure physics.) While it is true that science makes use of abstract terms such as wave functions and that physical terms are meaningful only within a theory, globally, physical theories are meaningful *only* in terms of observations.

The above applies to economics as well, where we are dealing with evolving complex systems. Once we define a unit of measurement or a basic measurement process, the entire economic theory is linked to this notion. Concepts such as inflation, growth, national product, and innovation are not a priori concepts but acquire meaning only through a process of measurement and observation linked to a theory. If we use an index such as the Laspeyres or Paasche index then inflation is defined through these indices.

A different approach to the problem of defining an index was proposed by Pia Nardini Malaney (1996), borrowing concepts from differential geometry. Differential geometry studies the properties of lines and curves in spaces of arbitrary dimensions using the tools of differential calculus. The reason why differential geometry might be a useful tool to our understanding of inflation is the following. At every moment, the competitive situation is re-evaluated and new prices are set. Given a basket of products, as the economy evolves prices describe a line in a space with as many dimensions as products.

Nardini Malaney observed that in order to compute an index we need to decide what remains constant through time. In the Laspeyres and Paasche indices, quantities are kept constant and new products are not introduced, but in reality quantities do change and new products are introduced. Nardini Malaney's reasoning is the following. At a given instant the economy is represented by a vector of quantities and prices: $v_{t_0} = (q_{t_0}, p_{t_0}), v \in R^{2n}$.

Suppose time moves from $t0$ to $t1$: $t0 \rightarrow t1, t \in [0,1]$. We can describe the path of the economy through a vector function: $\alpha(t) \rightarrow (\alpha 1(t),...,\alpha 2n\ (t))$ so that: $q_{i,t} = q_i\ (\alpha_i\ (t))$, $p_{i,t} = p_i\ (\alpha_i(t))$.

The output of the economy $v_t = (q_t,p_t)$ and its value $V_t \rightarrow q_t p_t$ change in time. However, at any instant, we can decompose any basket v_t into two baskets:

$$v_t = k v_{t0} + b_t$$

$$b_t = (b_1,...,b_{nt})$$

$$\sum p_{i_t} b_{i_t} = 0$$

where k is a scalar and bt is a basket of barters. Moving along the curve $\alpha(t)$ we

create a basket whose quantities are an exact multiple k of the initial basket. The value k is the value of the index. The process forming the baskets v_t, b_t can be described mathematically as parallel transport along the path $\alpha(t)$ keeping the *covariant derivative* equal to zero. Appendix 6.A introduces the notions of covariant derivative and parallel transport.

A comment is in order. Observe that in general indices are path-dependent. The evaluation of inflation or deflation over long periods of time depends on the paths the economy takes in moving from the initial to the final point. No direct comparisons over long distances in the product space are possible. Whatever index we use, we need to make comparisons over short periods of time, when changes are small. But the sequence of changes determines the final result.

How do we choose between the different approaches to measuring inflation? Any concept and measurement of inflation should respect essentially two conditions: (1) that it coincides with the true value of inflation under limit conditions such as all prices moving by the same multiple and (2) that it is useful in economic theory. The impossibility of defining a true index has serious consequences for economics. Here is why. The distinction between nominal and real quantities hinges on defining an inflation index, and the theory of growth depends on the distinction between real and nominal quantities. In addition, theories such as Monetarism can be discussed only in relation to a specific notion of inflation. Fisher established his index to give meaning to the Quantitative Theory of Money.

In a broader sense, we should define a notion of inflation fully integrated in economic theory. This is the method of science: a theory responds to the empirical test *in toto*. So we have to show that a quantitative measure of inflation is a useful concept. It is possible that inflation as currently measured by global indices is not a useful concept. We might ultimately need multiple inflation rates given that any single number fails to capture the complexity of modern economies.

6.4 Chartalism and the State Theory of Money

In modern monetary systems, there are essentially two ways to obtain *new* money: (1) as an individual or non-bank entity, borrow from banks and (2) sell assets such as stocks and bonds to the central bank using commercial banks as the intermediary. In countries where central banks can legally finance the state, one can also obtain new money by supplying goods or services to the state. Note that this rule does not apply to already existing money; already existing money can be obtained through exchange with other entities. Clearly this involves a transfer of money, not the creation of new money.

The supply of money from the state to the public is the topic of a theory called Chartalism (or Cartalism); the supply of money through bank credit is the topic of two competing theories called Verticalism and Horizontalism, both discussed in Chapter 5 in relation to the creation of money. Chartalism and Circuitism are also known as Verticalism and Horizontalism because Chartalism is a theory of the "vertical" flow of money from the state to the private sector while Circuitism is a theory of the "horizontal" flow of money inside the private sector.

The theory of Chartalism is generally attributed to Georg Friedrich Knapp. Knapp laid out the theory in *Staatliche Theorie des Geldes* (1905), which was translated into English only in 1924 under the title *The State Theory of Money*. At the time Knapp wrote his book, the prevailing view was that money must be something with intrinsic value, in practice gold or silver coins. Knapp coined the term "Metallism" to designate this theory, clearly opposed to the principle of Chartalism that he proposed in his work. Knapp had another challenger in the person of Carl Menger, who (as we saw in Chapter 3) developed the theory that money is not a "creature of the state" but instead was created by the private sector to optimize exchange.

Chartalism holds that money is always fiat money. Money has value because it is issued by the state which requires that taxes, fines, and other sums due to the state be paid in the money it issues. Chartalists argue that even if money is formed by coins it has value not because of the intrinsic value of the metal but because of the government seal on the coins. In the view of Chartalists, only the government has the power to make a specific money accepted. Knapp's views were not readily accepted. Moreover, the key role of fiat money, that is, of the state currency, has been weakened by the endogenous generation of credit money by commercial banks and, in many areas including the U.S., the U.K., and the euro area, by the fact that the central bank cannot directly finance the state.

Appendix 6.A An overview of differential geometry

Differential geometry can be studied at different levels of abstraction. The entry level is the study of surfaces in a Euclidean space. For the purpose of building an index of prices, we do not need the full power of differential geometry, only some basic notions and, in particular, the notions of covariant derivative and parallel transport, which we will now introduce.

Consider an open set U in R^2. A regular parametrization of a subset M c R^3 is a C^3 one-to-one function with nonzero derivatives: χ: $U \to M$, x_u x $x_v \neq 0$. A connected subset M c R^3 is called a surface if each point has a neighborhood that is regularly parametrized. If we fix u or v we obtain parametrized curves on M. At any point $P \in M$ the derivatives x_u, x_v are tangent to the u,v curves that pass through P.

The tangent plane of M in a point P, called T_pM, is the set spanned by the tangent vectors to the u,v curves in P. The *unit normal* to the parametrized surface is the vector: $n = \frac{x_u \times x_v}{\|x_u \times x_v\|}$. An important part ot differential geometry is in understanding the intrinsic geometry of a surface as perceived by a bidimensional observer on the surface. The parallel movement on the surface is one such property.

A *vector field* on is a vector function $X(P)$: $M \to R^3$ such that: $X(P) \in T_PM$, $\forall P \in M$, $X = (X,Y,Z)$, $P = (x,y,z)$.

For any parametrization $x(u,v)$: $U \to M$ the function:

$$X \circ x : U \to R^3$$

$$\begin{aligned} X \circ x &\equiv X\big(x(u,v), y(u,v), z(u,v)\big) \equiv \\ &\equiv X_1\big(x(u,v), y(u,v), z(u,v)\big)e_1 + X_2\big(x(u,v), y(u,v), z(u,v)\big)e_2 + \\ &\quad X_3\big(x(u,v), y(u,v), z(u,v)\big)e_3 \end{aligned}$$

$$e_1 = (1,0,0), e_2 = (0,1,0), e_3 = (0,0,1)$$

is a continuously differentiable vector function of the two variables u,v.

We can compute differentials and differentiate a vector field as usual along a parametric curve:

$$dX = dX_1 e_1 + dX_2 e_2 + dX_3 e_3 =$$
$$= \left[\left(\frac{\partial X_1}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial X_1}{\partial y}\frac{\partial y}{\partial u} + \frac{\partial X_1}{\partial z}\frac{\partial z}{\partial u} \right) du + \left(\frac{\partial X_1}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial X_1}{\partial y}\frac{\partial y}{\partial v} + \frac{\partial X_1}{\partial z}\frac{\partial z}{\partial v} \right) dv \right] e_1 +$$
$$+ \left[\left(\frac{\partial X_2}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial X_2}{\partial y}\frac{\partial y}{\partial u} + \frac{\partial X_2}{\partial z}\frac{\partial z}{\partial u} \right) du + \left(\frac{\partial X_2}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial X_2}{\partial y}\frac{\partial y}{\partial v} + \frac{\partial X_2}{\partial z}\frac{\partial z}{\partial v} \right) dv \right] e_2 +$$
$$+ \left[\left(\frac{\partial X_3}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial X_3}{\partial y}\frac{\partial y}{\partial u} + \frac{\partial X_3}{\partial z}\frac{\partial z}{\partial u} \right) du + \left(\frac{\partial X_3}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial X_3}{\partial y}\frac{\partial y}{\partial v} + \frac{\partial X_3}{\partial z}\frac{\partial z}{\partial v} \right) dv \right] e_3$$

Given any vector $V \in T_pM$, we choose a parametric curve $\alpha = (\alpha_1(t), \alpha_2(t), \alpha_{21}(t))$ with $\alpha(0) = P, \alpha'(0) = V$ and set: $D_vX = (X^\circ \alpha)'(0)$, or explicitly:

$$D_V X = (X \circ \alpha)'(0) =$$
$$= \left[\left(\frac{\partial X_1}{\partial \alpha_1}\frac{d\alpha_1}{dt} + \frac{\partial X_1}{\partial \alpha_2}\frac{d\alpha_2}{dt} + \frac{\partial X_1}{\partial \alpha_3}\frac{d\alpha_3}{dt} \right) e_1 + \right.$$
$$+ \left(\frac{\partial X_2}{\partial \alpha_1}\frac{d\alpha_1}{dt} + \frac{\partial X_2}{\partial \alpha_2}\frac{d\alpha_2}{dt} + \frac{\partial X_2}{\partial \alpha_3}\frac{d\alpha_3}{dt} \right) e_2 +$$
$$\left. + \left(\frac{\partial X_3}{\partial \alpha_1}\frac{d\alpha_1}{dt} + \frac{\partial X_3}{\partial \alpha_2}\frac{d\alpha_2}{dt} + \frac{\partial X_3}{\partial \alpha_3}\frac{d\alpha_3}{dt} \right) e_3 \right]_{t=0}$$

Given a vector field X and $V \in T_pM$, the *covariant derivative* is defined as the projection of D_vX on T_pM:

$\Delta_v X = D_v X - (D_v X n)n.$

Given a vector field X and a curve α, we say that the vector field X is *covariant constant* or *parallel* along α if:

$\delta_{\alpha'} X = 0$

These notions extend immediately to any number of variables. Consider an economy defined by a vector of quantities and prices: (q_t, p_t). Given any vector v, *consider* the following maps:

$$\Pi_{[q_t]}(v) = \left(\frac{p_t v}{p_t q_t} \right) q_t$$

$\pi_{\beta pt}(v) = v - \pi_{[qt]}(v)$

Suppose that σ represents the analog of the reference basket as the economy moves in time. The economic covariant derivative is:

$$\Delta_{\alpha}.\sigma = \Pi_{[q_t]} D_{\alpha}.\left(\Pi_{[q_t]}\sigma\right) + \Pi_{\beta_{p_t}} D_{\alpha}.\left(\Pi_{\beta_{p_t}}\sigma\right)$$

It can be demonstrated that if the covariant derivative is equal to zero then:

$\sigma_t = k_t q_t$

The scalar k_t is the index.

Notes

1 For example, the rounded-off figure for Japan is 40%, the U.K. 43%, Germany 44%, and Italy 50%. Source https://data.oecd.org/gga/general-government-spending.htm.

2 Source: http://data.worldbank.org/indicator/NE.CON.GOVT.ZS?locations=FR-DEIT-GB-JP-US.

3 Today, banks in the Federal Reserve System include the Federal Reserve Banks of Boston, New York Philadelphia, Richmond, Atlanta, Chicago, Saint Louis, Minneapolis, Kansas City, Dallas, and Sa Francisco.

4 For example, David Romer (2013), professor of political economy at the University of California-Berkeley counted six distinct shocks in the U.S. markets during the past 30 years or so; Joseph Stiglitz (2013) professor of economics at Columbia University and co-recipient of the 2001 Nobel Memorial Prize in Economic Sciences, counted approximately 100 financial crises worldwide in the past 30 years.

5 Recall from Section 1.2 that in the second millennium BC, King Seti in Egypt stipulated that the worth of worker engaged in the building of his temple was "'4 lb. bread, 2 bundles of vegetables and a roast of meat daily, and a clean linen garment twice a month'!" (Childe 1982 [1942], p. 131).

6 The ECU (European currency unit) was not a true currency but a basket of currencies used as an interna accounting unit for European Community (EC) members. It can be thought of as an earlier version of the euro though EC currencies participating in the ECU basket of currencies and the euro are not exactly the same. The value of the ECU/ dollar exchange rate at the time of the launch of the euro was 1.1686.

7 Purchasing power parity (PPP) is the notion that exchange rates should, over time, move towards rates such that baskets of the same products have the same value in different countries.

8 Source: www.economist.com/content/big-mac-index. *The Economist* article dated July 13, 2017 used a July 2017 exchange rate; the exchange rate at the writing of this book (August 2017) is effectively 0.88 euro to the USD.

9 "la croissance continue du niveau générale des prix," p. 55; the translation is the author's own.

10 Ragnar Frisch was a co-recipient of the first Nobel Memorial Prize in Economic Sciences in 1969 "for having developed and applied dynamic models for the analysis of economic processes." He is considered to be the founder of econometrics and was the first to use the terms macroeconomics and microeconomics. Money was, so to say, in his blood: his family worked for some 300 years with precious metals including silver and gold.

References

Aftalion, Florin and Patrice Poncet. 1995. *Le Monétarisme*. Paris: Presses Universitaires de France.

Bridgman, Percy. 1927. *The Logic of Modern Physics*. New York: Macmillan.

Childe, Gordon. 1982 [1942]. *What Happened in History*. London: Penguin Books.

Frisch, Ragnar. 1930. "Necessary and Sufficient Conditions Regarding the Form of an Index Number Which Shall Meet Certain of Fisher's Tests." *Journal of the American Statistical Association*, vol. 25, no. 172 (December): 397–406.

Graziani, Augusto. 1990. "The Theory of the Monetary Circuit." *Economies et Societies*, vol. 24, no. 6:7–36.

Investment Company Institute. 2016. *2016 Investment Company Fact Book: A Review of Trends and Activities in the U.S. Investment Company Industry*. 56th ed. www.ici.org/pdf/2016_factbook.pdf.

Knapp, Georg Friedrich. 1924. *The State Theory of Money*. London: Macmillan and Co.

Konüs, Alexander A. 1939. "The Problem of the True Index of the Cost of Living." *Econometrica* 7 (January):10–29.

Menger, Carl. 1892. "On the Origins of Money." *Economic Journal* 2:239–255; translation is by C.A. Foley.

Milana, Carlo. 2009. "Solving the Index-Number Problem in a Historical Perspective." EU KLEMS Working Paper No. 43.

Mill, John Stuart. 1852. *Principles of the Political Economy*. London: John W. Parker & Son.

Mishkin, Frederic S. 2007. *The Economics of Money, Banking, and Financial Markets*. 8th ed. Cambridge, UK: Pearson.

Nardini Malaney, Pia. 1996. "The Index Number Problem: A Differential Geometric Approach." Harvard PhD Thesis.

Popper, Nathaniel. 2012. "Stock Trading Is Still Falling after '08 Crisis." *New York Times*. (May 6): www.nytimes.com/2012/05/07/business/stock-trading-remains-in-a-slide-after-08-crisis.html.

Romer, David. 1996. *Advanced Macroeconomics*. New York: McGraw-Hill.

Romer, David. 2013. "Preventing the Next Catastrophe: Where Do We Stand?" Presentation at the IMF conference "Rethinking Macro Policy II: First Step and Early Lessons." (May 9).

Samuelson, Paul. 1976. *Economics*. London: McGraw-Hill, International Student Edition.

Stiglitz, Joseph. 2013. “The Lessons of the North Atlantic Crisis for Economic Theory and Policy.” *IMFBlog*. (May 9): https://blogs.imf.org/2013/05/03/the-lessons-of-the-north-atlantic-crisis-for-economic-theory-and-policy/

Trevithick, James. 1977. *Inflation*. London: Penguin Books.

Von Mises, Ludwig. 1912. *Theorie des Geldes und der Umlaufsmittel*. 1st English version published 1934 under the title *Credit and Money*. London: Jonathan Cape Ltd.

7
Money

How it's distributed

7.1 The question of the distribution of money

In previous chapters we discussed the nature and generation of the four forms of money typical of modern developed countries: coins which are minted; banknotes which are printed by central banks; bank deposits – ultimately a record in a computer; and reserves which are commercial banks' deposits with the central bank, also created as records in a computer and no longer strictly linked to base money. The costs of money generation have shifted from the cost of precious metals to produce coins to the cost of a secure computer and communications infrastructure. Let's now turn our attention to the distribution of new money and in particular how distribution has evolved and possible future evolutionary paths.

Let's separate the conceptual building blocks. Recall that the creation of today's fiat money or bank deposits is in itself a simple process; what is important is how this newly created money is allocated. For example, the creation (or extension) of bank deposits in itself consists in writing a number in a computer file. From the point of view of a user of the banking system, what matters is the number that keeps track of the "true" value of the deposit(s). How the deposit was created is known only to the bank and the owner of the deposit. When a client receives a payment, be it a check or a bank transfer, what is important is that the payment is accepted and the due amount correctly credited to his/her account.

Looking at the process of the allocation of money from an engineering point of view, there are a number of conditions that the allocation of money should respect. First money should be allocated to those entities that can make an efficient use of money. But even more important is that the allocation of money should respect the notion of justice.[1] In order to maintain trust in a system that generates money, the allocation of new money must be subject to transparent, socially acceptable rules. The allocation of new money is a critical function in modern economies: it is responsible for the dynamics of money. Conceptually, various mechanisms would allow a system that generates money to "inject the money into the real economy," that is to say, to distribute it.

A fresh look at the question of the allocation of money is discussed in some recent books and papers. Among them, McLeay et al. (2014a, 2014b) and Ryan-Collins et al. (2011) describe the process of endogenous money generation and allocation as currently implemented by central banks; Wray (2012), Keen (2011), and Mosler (2010) discuss Modern Money Theory; Niall Ferguson (2008) and Felix Martin (2013) present the evolution of the concept and use of money. Classical textbooks on banking often give a somewhat distorted picture as they still adopt a theory of banks

as intermediaries. Werner (2016) presents a discussion of the different theories of the banking system and offers some empirical evidence that banks are creators of money without needing to be intermediaries, that is to say, banks create money *ex nihilo* without the need to receive deposits.

Note explicitly that we are discussing how *newly created* money is allocated to economic agents. Money that already exists in the system circulates and can be obtained as the result of business transactions, the payment of wages, reimbursement by the state, or a myriad other legal or private motives for receiving a payment. Once money has been created and injected into the economy, the rules that make money circulate are the business or legal rules that justify payments.

Let's go back to the original question: Who gets the newly created money? What is the meaning of expressions such as "injecting money into the economy"? With money we can buy things, get things done. So, who gets the money? Who gets the power that goes with money?

7.1.1 The allocation of money

Recall from Chapter 2 that in modern developed economies money can be coins, banknotes, or bank deposits but the definition of money is somewhat arbitrary as other very liquid assets can be used as storage of value and converted into money at par on demand. We use a definition of money that includes coins, banknotes, and bank deposits. Reserves are a form of money used only between central banks and commercial banks in the national banking system. New money can be legally obtained in different ways in function of the type of money. New bank deposits can be legally obtained in two ways:

- by selling assets to the banking system;
- by taking out a loan with a commercial bank;

and new coins or banknotes also in two ways:

- by buying them with bank deposits;
- by (in some countries) providing services to the government.

The latter is a controversial point. While in some countries, for example India, the government is allowed to borrow directly from the nation's central bank and banknotes can thus be obtained directly in exchange for supplying goods or services to the state, governments in the U.K., the U.S., and the euro zone cannot receive

money from their respective central banks; they must therefore pay for goods and services with money obtained from taxation or debt, which is money already in the hands of the public. The public, therefore, cannot obtain *new* money from the state. The financing of the state is subordinated to the generation of money by private entities such as commercial banks.

In modern monetary systems, the role of base money, that is, hard currencies such as coins and banknotes, is being progressively reduced. For example, in discussing the English banking system, Ryan-Collins et al. (2011) write: "under normal conditions even cash, which is physically 'created' by the Bank of England, cannot enter circulation without customers first exchanging it for commercial bank money that they already hold" (p. 117, 2016 edition). The authors recall that the Treaty of Maastricht, effective as of November 1993, created the European Union and led to the creation of the European Central Bank (ECB). They note: "One of the most important rules of the Treaty in relation to money creation is Article 101 EC, now known as Article 123* of the Treaty of the Functioning of the European Union (TFEU). This article prohibits the direct financing of government spending by the nation's central bank" (p. 118, 2016 edition).

The key point in modern economies where the central banks cannot finance the state is the following:

> *Newly created money (bank deposits) can be obtained either by selling assets to the banking system or by taking out a loan. Newly created coins or banknotes can be obtained (i.e. purchased) only via bank deposits.*

Above all, newly created money cannot be distributed for private gain. The legal framework of money today is linked to the legal framework of ownership. Private ownership is reinforced by the law; money follows as a tool to transfer ownership and to store value for transferring ownership in the future. No private entity can issue money with the sole objective of acquiring the ownership of goods, assets, or the fruition of services. In the current system, new money can be issued and distributed only under the promise that it will be returned, that is, that it can only be lent. The only exception is when central banks purchase assets using banknotes or crediting bank deposits. Note that alternative money created by private entities such as bitcoins is paid for with banknotes or debiting bank deposits.

7.1.2 Assets and liabilities

An important point related to the generation and distribution of money that needs to be clarified is the following: money and financial assets come in pairs, assets and

liabilities. Someone's assets are someone else's liabilities. The sum is zero. Wray (2015) writes: "It is a fundamental principle of accounting that for every financial asset there is an equal and offsetting financial liability. The checking deposit . . . is a household's financial asset, offset by the bank's liability (or IOU)" (p. 1).

Note, however, that this asset/liability pair is related to accounting conventions: the fact that for each financial asset there is a corresponding liability does not mean that all financial assets are instances of a credit/debt relationship. Textbooks define financial assets as contracts that give the owner of the contract the right to receive a stream of future payments. A bond, for example, gives the bondholder the right to receive coupons and repayment of the principal. However, when we go back in the chain of assets to fiat money, the notion that for "every financial asset there is an equal and offsetting financial liability" is an accounting convention, not a true credit/debt relationship. Banknotes in circulation appear as liabilities in the balance sheet of the central banks, but clearly this does not mean that when central banks issue banknotes they incur a debt. Again, the fact that banknotes appear as liabilities is an accounting convention (see Chapter 4).

An owner of banknotes considers the notes an asset because he or she can use them to purchase goods and services; the central bank considers banknotes a liability by an accounting convention. Today, a bank account is a liability (or debt) of a commercial bank in that the holder of the account can ask the bank to deliver banknotes to extinguish the account. But if banknotes were to be eliminated, bank accounts could still be considered liabilities in a formal sense though they would no longer be considered a debt. In the absence of banknotes, a bank could only be asked to transfer a client's account to another bank.

The above is a clear instance of the importance of the operational approach to defining concepts and variables. What do we really mean when we say that a banknote is an IOU? Before states went off the gold standard, for example, it meant: I Owe You the equivalent in gold. But today, operationally speaking, it does not mean that the central bank has a debt. Viewing the situation from an operational point of view, it is clear that today's system for the generation and distribution of money can change in response to technological and/or social changes. In fact, the current way of creating and distributing money is an engineering choice, not a law of nature. Changes might include the decoupling of bank deposits and loans, the introduction of a basic income, or helicopter money as discussed later in this chapter.

In aggregate, new money is a tool borrowed to perform transactions, be it for consumption or investments (see Chapter 8 on Circuitism for a theoretical formulation of this principle). Today, in aggregate, the bulk of money needed by individuals or firms is obtained through bank loans. Issuing bonds or shares might

seem to be a different form of financing, but bonds and shares are bought by entities that already have money. There has, nevertheless, been a shift in borrowing: in the last three decades firms have reduced their borrowing while households have progressively increased theirs. Private individuals typically take out loans to purchase big-ticket items such as houses, automobiles, and education for their children and, increasingly, to finance the purchase of low-ticket items including consumer goods. The dynamics of stagnant salaries, household borrowing, and increased corporate profits are discussed in Chapter 8 together with the data. All money obtained through banks has a limited lifespan as loans must be repaid and money therefore destroyed.

7.1.3 Financing the state

A critical issue from the point of view of fiscal policies is whether central banks can create money and distribute it to the government to pay for the cost of government (defense, debt financing, administration, health and education, etc.). Stated differently, can the government itself create all the money it needs to finance expenditures? (See Chapter 8 for a discussion on the financing of government operations and the question of national debt from the point of view of Modern Money Theory.) Jácome et al. (2012) summarize international practices regarding the monetization of government debt:

> (i) in most advanced countries, central banks do not finance government expenditure; (ii) in a large number of emerging and developing countries, short-term financing is allowed in order to smooth out tax revenue fluctuations; (iii) in most countries, the terms and conditions of these loans are typically established by law, . . . and (iv) in the vast majority of countries, financing other areas of the state, such as provincial governments and public enterprises, is not allowed.
>
> *(p. 1)*

Let's now look more closely at the distribution of different forms of money.

7.1.4 The distribution of coins

The framework relative to the distribution of coins and banknotes described in the above paragraphs – and according to which coins and banknotes are sold to the public in exchange for bank deposits – did not always apply in the past. Coins could be freely minted by any holder of gold or silver. There were of course various ways to obtain the precious metals: through princely monopolies on the sources, private monopolies obtained from rulers to exploit mines, victorious generals who took the

metals as booty, or through commercial activity such as long-distance trade.

But lack of gold or silver was always a threat to governments and economies. A prince with no gold or silver could not mint coins (so would likely have problems paying his soldiers). Solutions included debasement or the accumulation of debt. Luckily for today's governments (and economies) coins are now used only for small transactions and have no intrinsic value even if, relative to their value, they are expensive to produce.

7.1.5 The allocation of banknotes

With banknotes things change considerably with respect to coins: production is rigorously controlled, costs limited, and, importantly, there are virtually no physical constraints to the number of banknotes that can be printed. When banknotes were first issued and distributed in Europe at the end of the 17th century, first in Sweden and then in England with the creation of the Bank of England (BoE) in 1694, they were essentially a replacement for gold or silver coins, a sort of personalized receipt to acknowledge the deposit of the metal coins and guarantee their convertibility back into gold or silver on demand.

Not so long ago commercial banks also had the right to issue banknotes. Newly printed banknotes were distributed either in exchange for gold or silver coins or when taking out a loan. Notes issued by commercial banks were different in their nature and function from today's state-issued banknotes. In fact, they were a debt of a bank redeemable in gold or silver or in the state currency.

In the second phase, starting in the early 20th century, banknotes were to be progressively delinked from gold or silver or any other good or commodity. The move to break the link between banknotes and a fixed value in gold (the so-called gold standard) started in the U.K. in 1931; the U.S. was to abandon the gold standard 40 years later, in 1971. Banknotes had become fiat money.

Banknotes are the embodiment of power: an entity that has the power to create banknotes and to enforce their status as fiat money has power, not only economic power. Thus the production and distribution of banknotes as fiat money have always been kept strictly in the hands of governments.

The mechanics of the distribution of banknotes in most developed countries is the following. Once printed, banknotes are delivered to the central bank. As long as they stay in central banks' vaults, banknotes are simply nicely colored pieces of whatever support they are printed on. On demand, central banks physically deliver the banknotes to commercial banks and debit their reserve accounts or (exceptionally)

might lend banknotes to commercial banks.[2] Commercial banks, in turn, give, on demand, banknotes to their clients (typically via ATMs for small amounts) and debit their accounts. We can say that commercial banks "sell" banknotes to their clients, debiting their accounts.

Banknotes are a permanent form of money. Once in the hands of the public, they are not returned to the central bank. The distribution of banknotes must therefore be a sale, not a loan as loans have to be repaid, thereby destroying the money. Exhibit 7.1 illustrates the evolution of the value of banknotes in circulation in the U.S.

After leaving the central banks' vaults for commercial banks' vaults to be made available to the public, banknotes appear as liabilities in the balance sheet of central banks. This accounting practice is currently followed by all central banks. The Bank of Japan, for example, in the FAQs section of its website, has an entry for the question: "Why are banknotes on the liability side of the Bank's balance sheet?" The answer is that historically speaking, banknotes were a liability because the holder could demand their convertibility into gold or silver. However, even though the gold standard has been abandoned, "banknotes continue to represent the liability certificate – of which the credibility must be ensured by the Bank – and are still on the liability side of the Bank's balance sheet. Such accounting treatment of banknotes is commonly applied among other major central banks."[3] That is, banknotes are on the liability side to make sure that the commitment of the bank to financial stability is reinforced.

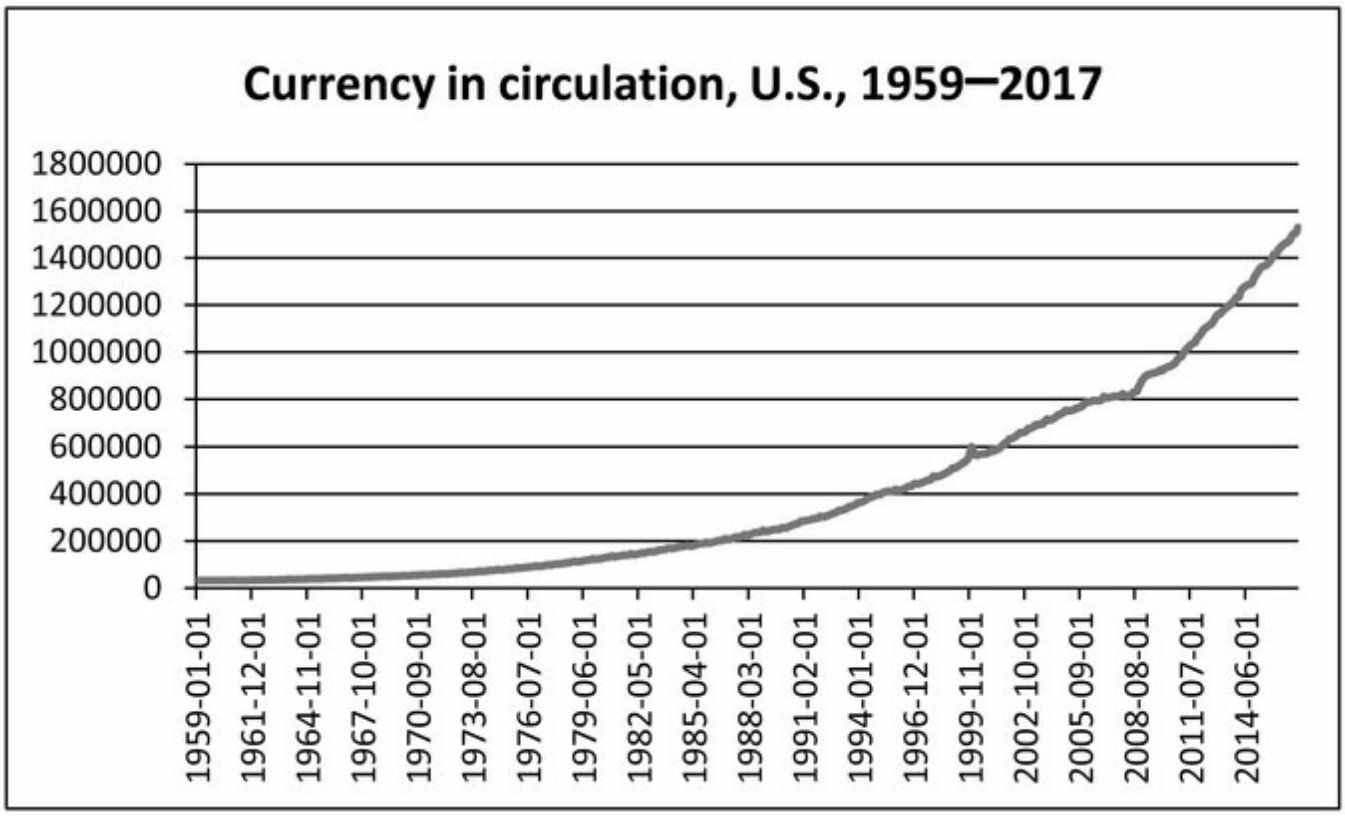

EXHIBIT 7.1 The value of banknotes in circulation in the U.S. from 1959 to 2017 in millions of dollars. Constructed by the author using data obtained from the FRED (Federal Reserve Economic Data) of the Federal Reserve Bank of Saint Louis.

There might be arcane technical reasons why, in the highly unlikely case of the failure of a central bank, each bearer of banknotes has a claim on the real assets of the central bank. But this is not what is meant by saying that banknotes are liabilities of the bank. As stated by the Bank of Japan, putting banknotes on the liability side is an accounting convention that gives credibility to the issuing of banknotes.

7.1.6 The allocation of reserves

Reserves are a type of money – a deposit or an entry in the computer system – at a central bank in favor of commercial banks within its banking system. They are created by a central bank when it buys assets from banks in the national banking system and credits their reserve accounts. Only the central bank can create the reserves and change their value. In normal operations, a central bank buys a commercial bank's assets in the form of treasury notes, but under nonconventional operations such as quantitative easing (QE), central banks also buy commercial bonds and stocks from non-bank financial entities using the commercial banks as intermediaries. The effect of QE on bank reserves is illustrated in Exhibits 7.4–7.7 later in this chapter.

Box 7.1 The Changing Nature and Functions of Money

In modern advanced economies money is a tool that is *borrowed* to accomplish economic objectives such as to purchase goods and services or to organize production. Money has lost its character of a permanent storage of wealth to become a temporary tool that gives its holders the power to implement their economic decisions.

The purchasing power represented by money is allocated to different economic agents by commercial banks. Commercial banks have become the manufacturers and the distributors of money. The role of the state is reduced to controlling the conditions, that is, the price, at which money is supplied and to providing the base money to anchor the value of money.

Base money, formed by coins and banknotes, has a reduced role in modern economies as it represents only a tiny fraction (roughly 3–10%) of the money in circulation and is not necessary to generate the broad money formed by bank deposits. However, base money is still important: it provides the basic denominator of credit money. Today, the value of a bank deposit is measured in units of the base currency: dollars, euros, renminbi, yen, and other. In principle, bank deposits can be redeemed into banknotes either on demand or with some delay. Should banknotes be abolished, base money would need to be replaced by some other mechanism to fix the value of deposits.

7.1.7 The "distribution" (opening) of bank deposits

In modern advanced economies, banknotes represent only a small fraction of the total money in circulation. Exhibit 7.2 illustrates the evolution of the quantity of currency (banknotes and coins) and the quantity of aggregate money (M2 = banknotes, coins, and deposits) in circulation in the U.S. for the 53.5-year period from January 1959 to end-July 2013.

Exhibit 7.3 illustrates the ratio of the currency component (M1) and M2 in the U.S., again for the 53.5-year period from January 1959 to end-July 2013. As can be seen, the currency component remains small, oscillating in the range 0.06–0.12%.

Compare the allocation of bank reserves with the allocation of bank deposits. Reserves are bank deposits with the central bank, created, as we have seen, when a commercial bank borrows reserves from the central bank or when a commercial bank sells assets to the central bank. In both cases the central bank creates new

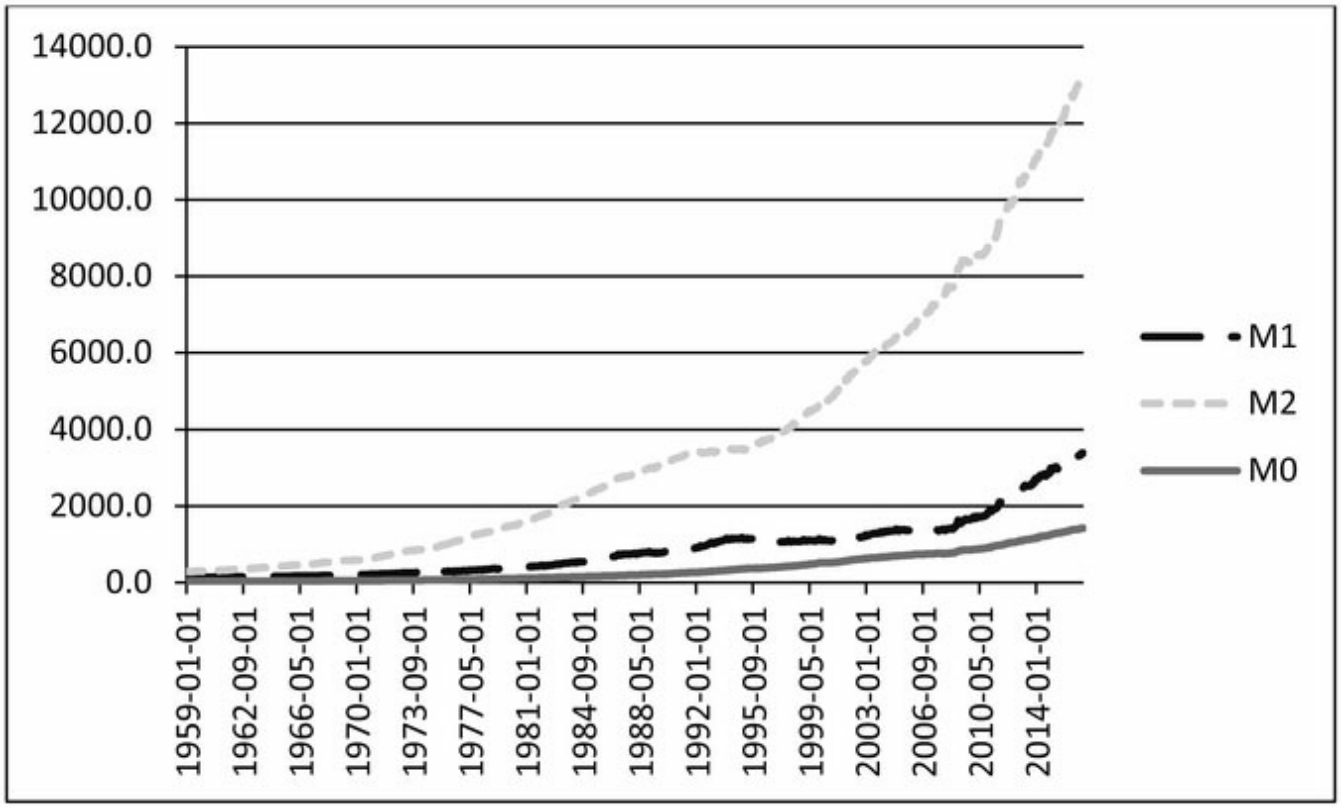

EXHIBIT 7.2 The evolution of the quantity of currency: M0 (banknotes and coins), M1 (M0 plus demand deposits), and M2 (M1 plus time deposits) for the 53.5-year period from January 1959 to end-July 2013 in millions of dollars. Constructed by the author using data obtained from the FRED (Federal Reserve Economic Data) of the Federal Reserve Bank of Saint Louis.

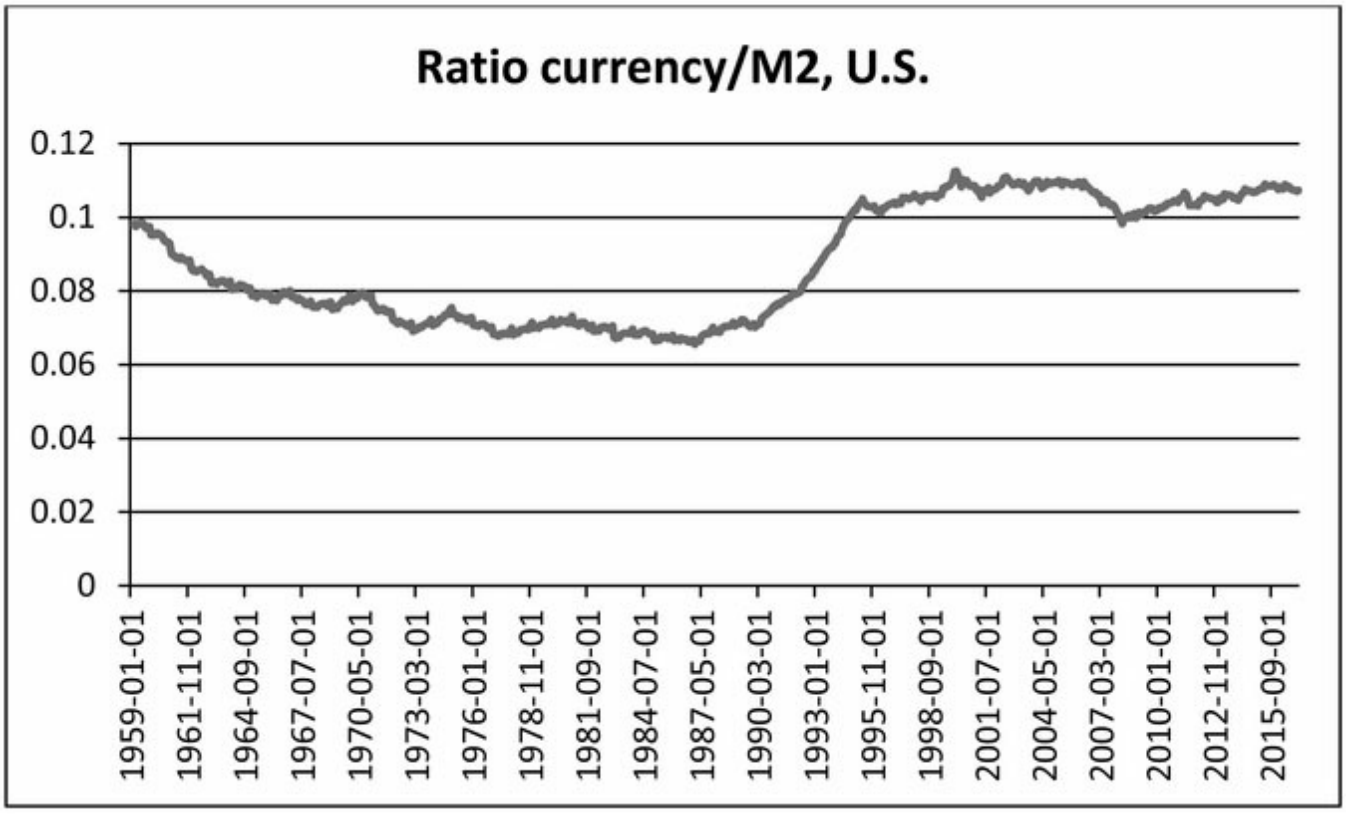

EXHIBIT 7.3 The ratio of the currency components M1 and M2 in the U.S. for the 53.5-year period from

January 1959 to end-July 2013. Constructed by the author using data obtained from the FRED (Federal Reserve Economic Data) of the Federal Reserve Bank of Saint Louis.

reserves *ex nihilo*. New bank deposits are created when a client borrows money from a commercial bank or when a bank purchases assets from a client. Even in the latter case, new bank deposits are created *ex nihilo*.

In both cases the balance of the operations is in principle zero. If a central bank lends reserves, it writes the loan as an asset and the reserves as a liability. In the same way, a commercial bank writes a loan as an asset and the relative deposit as a liability. Should the client move the deposit to another bank, the bank that made the loan would lose a deposit, which is a liability, but it would also lose a corresponding amount of reserves, which are an asset, summing to zero.

Bank deposits can be thought of as storage of creditworthiness, used to purchase goods or services. Convertibility of bank deposits into banknotes is not necessary to the functioning of bank deposits as money. However, in the absence of convertibility should banknotes be abandoned, both deposits and reserves would be considered liabilities by accounting convention; they are not debt.

Box 7.2 The Distribution of Money in a Nutshell

Money is necessary for the functioning of modern market economies. Currently there are four forms of money in advanced economies:

- coins;
- banknotes;
- reserves;
- deposits.

All four forms of money can be created and the newly created money added to the existing stock of money. Coins can be minted, banknotes printed, reserves created as records in central banks' computers, and deposits created as records in commercial banks' computers.

How is new money injected into the economy? How does it reach economic agents?

Coins can be obtained in exchange for banknotes or bank deposits. However, their global value is small: in practice, coins are used for small daily transactions, such as at the bakery.

Banknotes are a permanent form of money. In principle, once in the public's

hands banknotes remain in circulation indefinitely. Banknotes enter into circulation through banks with a two-step process. Step 1, banknotes are delivered to commercial banks by the central bank in exchange for reserves, that is, the central bank sells banknotes to commercial banks. Occasionally, banknotes can also be lent by the central bank to commercial banks. Step 2, commercial banks, in turn, distribute banknotes to the public, individuals, and entities, in exchange for deposits. Though together with coins banknotes form the base money, no private individual or entity can obtain banknotes directly from the central bank.

Reserves are a form of money used only for bank-to-bank relationships. Reserves are the accounts of commercial banks with their central bank. They are created by the central bank and attributed to commercial banks in exchange for assets.

Deposits are accounts attributed to an individual or an entity such as a firm by commercial banks after the holder of the account deposits an equivalent of legal tender (in which case no new money is created) or when the holder of the account takes out a loan (in which case new money is effectively created). Alternatively, new deposits are created when a central bank buys assets from a non-bank organization via an intermediary bank.

There are two circuits for the creation and distribution of money. One circuit – the one that provides the bulk of new money – is formed by commercial banks in creating bank deposits. Money obtained as a loan through the commercial banking system is not permanent money: it is generated when loans are taken out and destroyed when loans are repaid, in a self-sufficient process that does not require any exogenous input. The other circuit is formed of money obtained through the central banks. Central banks issue banknotes and create new reserves when they buy assets from the commercial banks.

In the future, central bank digital currencies might be added to the monetary system, taking over the role of banknotes or becoming a new type of money possibly with allocation rules different from the allocation rules of bank deposits based on loans.

7.2 Who gets the (new) money?

Let's now summarize the ways in which new money is allocated, starting with conventional ways to distribute money followed by non-conventional ways to distribute money.

7.2.1 Conventional ways to distribute (new) money

Who gets the new money? That is, who gets the ability to purchase goods and services that comes with the creation of new money, that is, new bank deposits?[4]

The answer is simple: those who take out loans or those who supply assets to the banking system get the new money.[5] New money created by commercial banks is formed only when an individual or an entity takes out a loan – with one exception: when the banking system buys financial assets such as bonds or stocks from non-banks, thereby creating money without corresponding loans. There might be other marginal circumstances that result in an imbalance between money and loans. If, for whatever reason, a client does not repay a loan, the bank has to cancel the debt. In so doing, however, the bank experiences a loss that must be compensated by a reduction of the bank's capital which is ultimately formed by financial assets.

Are there other ways to distribute money? The system currently in use in advanced economies does not allow any private organization or individual to print money or to appropriate money created by the central bank. Any new money therefore corresponds to a new loan. Only the government, in some but not all economies, can appropriate the new money and pay for goods and services with this money.

Therefore, when talking about money as credit, a distinction has to be made between deposits which are credit of individuals and entities such as firms with respect to the banking system, and loans that individuals or firms incur in order to acquire deposits, that is, purchasing power.

To summarize, in modern economies where central banks cannot supply money directly to their respective governments, money in the form of banknotes or bank credit can be obtained, in aggregate, either by selling assets to the central bank if a commercial bank or to a commercial bank if a private entity or by taking out a loan. There is no other way for money to enter the economy.

7.2.2 Unconventional ways to distribute (new) money: QE, QEP, HM, FM TDBs

Governments and central banks wanting to kick-start stagnant economies, increase consumers' purchasing power, or mitigate the effects of high unemployment have a number of policies they can implement and among the first of these is targeting interest rates. A reduction of interest rates should, in theory, propagate throughout the economy, reducing the cost of financing investments for firms and thereby increasing the demand for loans. An increase in loans outstanding, as we have seen, increases the quantity of deposits and therefore the quantity of money in the economy. More jobs and stronger demand should do the rest.

The conventional tools to influence rates are open market operations by which central banks trade short-maturity treasury bonds, either buying them from or selling them to commercial banks. In practice, central banks use methods different from simply buying or selling assets but the final result is the same.

Unconventional ways of distributing (new) money: quantitative easing

When in the post-2008 financial crisis period, interest rates reached their effective lower limits (close to zero or even negative) as is/was the case in many developed countries, central banks turned to unconventional monetary tools. The first among these to be employed was quantitative easing (QE), a term coined by the Bank of Japan who first implemented QE in 1990 in an operation consisting in buying assets other than short-term treasury bonds from non-bank entities. By purchasing assets such as long-maturities bonds and (in the case of the Bank of Japan[6]) corporate stocks from non-bank financial entities such as insurers, pension funds, and asset managers, central banks inject money directly into the economy.

Here is how QE works. Suppose that a non-bank financial institution sells bonds or other assets to its respective central bank. The transaction will not be done directly but through the institution's corresponding commercial bank. In the end, the non-bank financial institution has fewer bonds but more cash in its portfolio and the institution's bank thus finds itself with a corresponding increase in reserves. This cash is newly generated directly by the central bank without the need for a loan to be created by a commercial bank. Therefore, cash created with QE will not be destroyed as with the repaying of loans, but stays in the economy until the central bank decides to reverse QE, selling securities.

The objective of QE is to encourage a rebalancing of portfolios. In fact, those entities that sold their assets to the central bank find themselves with excess cash and will therefore try to rebalance their portfolios by acquiring assets that yield higher returns than cash, for example corporate shares. The result is that asset prices go up,

making more cash available to (in theory) support corporate investments. In practice QE has taken different forms in different countries and to varying effect, but in general it has been criticized for having undesirable distributional consequences, putting more money in the hands of those who already have it and failing, therefore, to encourage spending.

Unconventional ways of distributing (new) money: helicopter money, fiscal money, and free money

There are other expedients to increasing the money supply to support a stagnant economy including unconventional fiscal policy, though for the most part these policies have not gone further than the discussion phase. One such possibility is "helicopter money" (HM), so named after the U.S. economist Milton Friedman (1969) suggested, in his essay "The Optimum Quantity of Money," the possibility of a helicopter dropping $1,000 in bills from the sky in an attempt to put purchasing power directly into people's pockets. Such a policy has been referred to as Quantitative Easing for the People (QEP). The idea is not so dissimilar from experiments with local currencies inspired by the Swiss economist Silvio Gesell and discussed in Chapter 3. An important element in Gesell's local currency was "stamped money" to ensure that the money be spent rapidly as unspent balances were taxed.

A group of Italian scholars recently proposed what they call Fiscal Money (FM) as the best way to overcome the liquidity trap and increase consumption, distributing what they call the Tax Discount Bonds (TDBs) ("free money") over a period of three years, to a total equal to 2–3% of national GDP. Their idea: the TDBs would be allocated to families in inverse proportion to income and assigned to companies in proportion to the number of their employees, thereby reducing the cost of labor and increasing firms' competitiveness in an effort to retain a balance in trade. According to the proposal, the state would also be able to use TDBs to pay for public works and increase employment and private investment.

Here, according to Enrico Grazzini (2016) is how TDBs would work: TDB would be euro-denominated bonds issued by a national government and valid for tax discount, maturing 2–3 years from issuance, but negotiable on financial markets and so immediately convertible into legal currency. An argument advanced in favor of TDBs: the assumption that the fiscal multiplier exceeds 1 (one) when capital and labor resources are greatly underutilized and interest rates are close to zero as is now the case in the euro zone.

7.3 Persuading the sceptics: loans and bank deposits

The notion that new money is allocated by commercial banks in according loans as loan–deposit pairs might seem counterintuitive, as might the notion that money is destroyed when debts are repaid. Let's look at empirical data. Consider the U.S. economy and compare the total amount of loans and leases with the total amount of deposits. Exhibit 7.4 shows the total loans and leases in the U.S. over the 41.5-year period between January 1973 and mid-2015 and the total deposits at U.S. commercial banks over the same period. Note that total deposits are not the same quantity as M2 which, in addition to deposits, includes currencies and money market mutual funds.

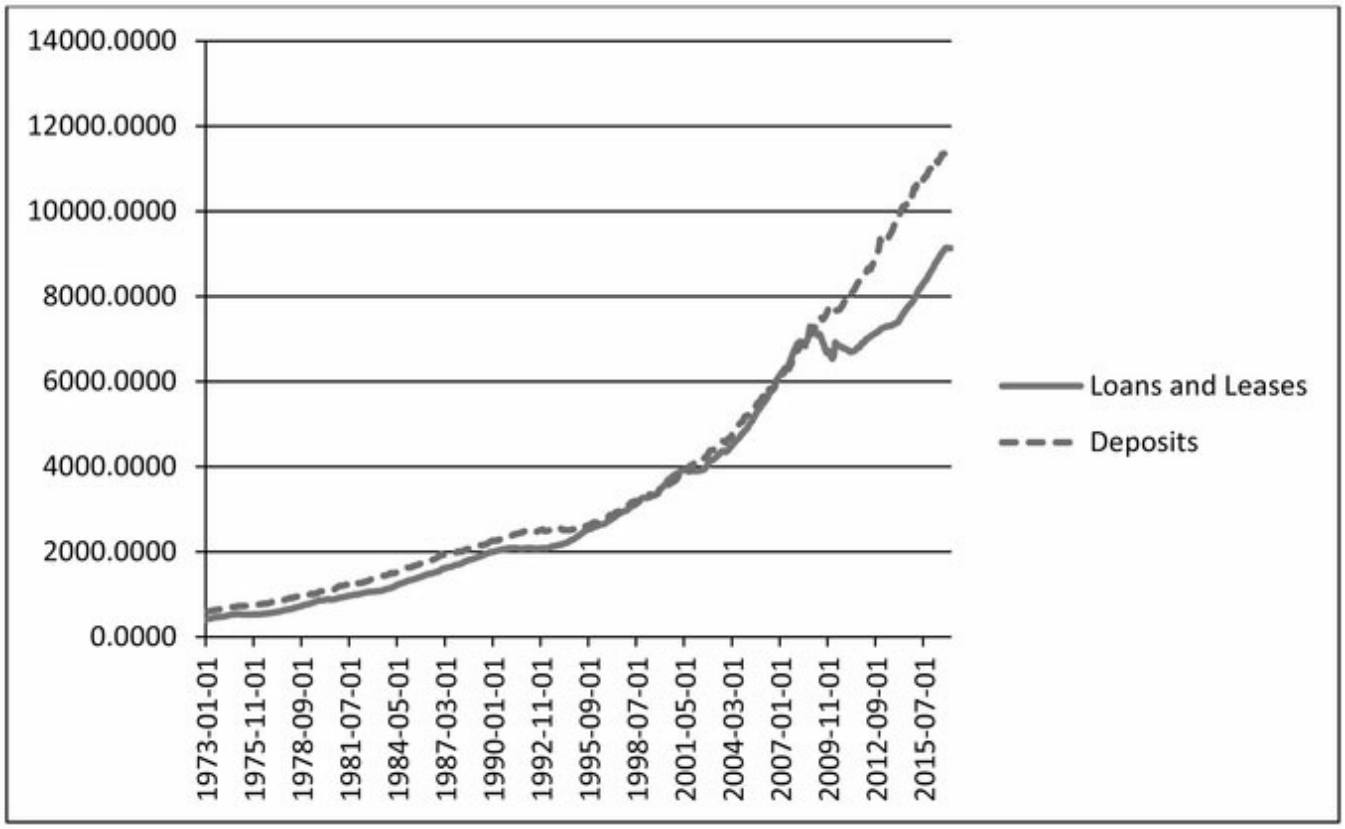

EXHIBIT 7.4 The total loans and leases and the total deposits at U.S. commercial banks over the 41.5-year period from January 1973 to mid-year 2015 in billions of dollars. Constructed by the author using data obtained from the FRED (Federal Reserve Economic Data) of the Federal Reserve Bank of Saint Louis.

As we can see from Exhibit 7.4, the total amount of deposits was slightly higher than the total amount of loans and leases up to the year 2008, after which the amount of deposits significantly increased vis-à-vis loans and leases. The difference up to the year 2008 is explained primarily by the open market operations[7] through which central banks purchase short-term Treasury notes from banks. The significant

difference after 2008 is explained by QE as will be seen in the following exhibits.

In October 2008, the U.S. Federal Reserve started its QE program. Under th program, the Fed bought Treasury securities of different maturities and mortgage-backed securities. Exhibit 7.5 illustrates the progression of purchases of the two types of assets for the 14-year period from December 18, 2002 to December 18, 2016.

As can be seen from Exhibit 7.5, the Fed held approximately $600 billion in Treasury notes in 2007; through the QE program, by end-2014 the Fed held almost $2.5 trillion in Treasury notes, an increase of more than 400%. Before 2008, the Fed held no mortgage-backed securities; through the QE program, by end-2014 the Fed held almost $1.8 trillion in mortgage-backed securities.

The massive program of QE explains why, starting in 2008, the curve of deposits runs considerably higher than the curve of loans and leases: the Fed provided deposits without loans. Exhibit 7.6 illustrates the evolution of reserves at U.S. commercial banks. Limited until 2009, reserves grew progressively to reach almost $2.8 trillion by 2014.

If we add reserves to loans and leases, we obtain a graph that follows quite closely the graph of deposits, as can be seen in Exhibit 7.7.

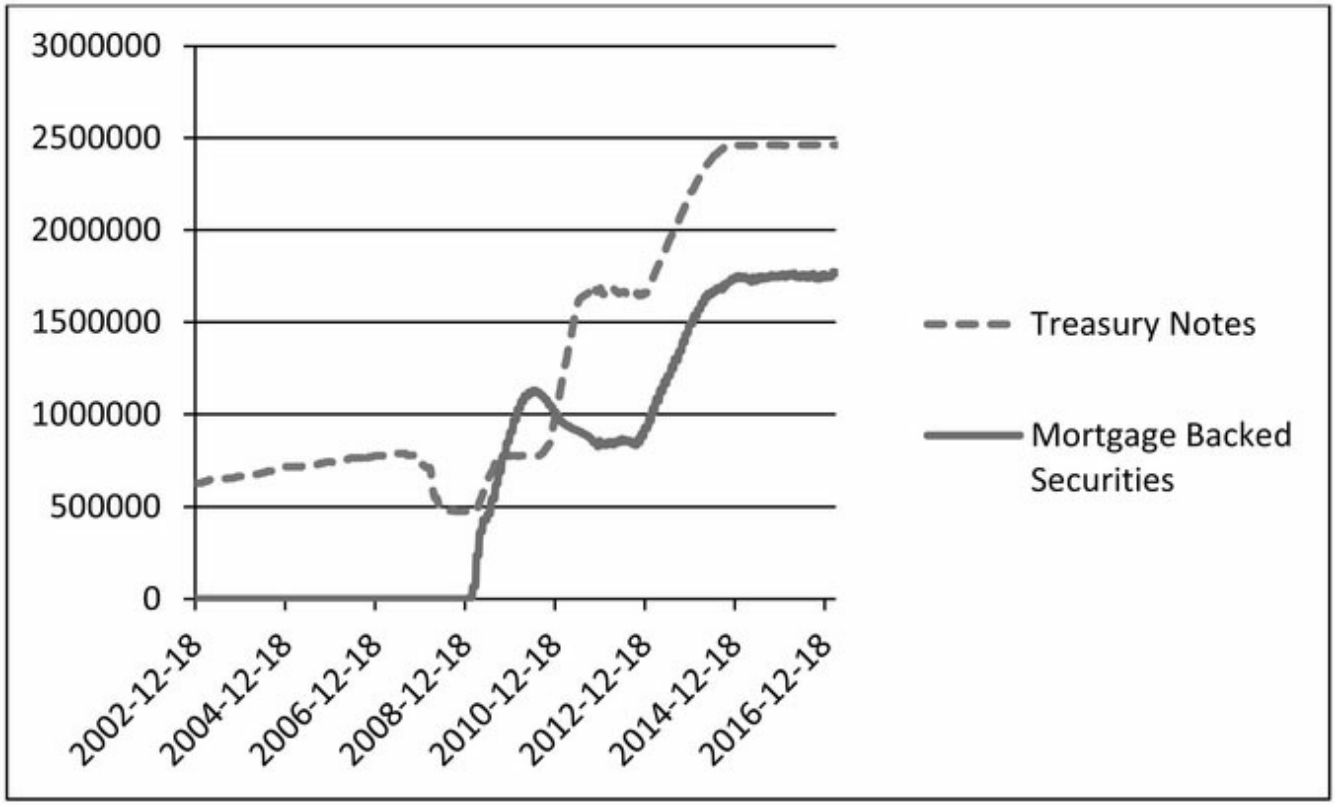

EXHIBIT 7.5 The evolution of purchases by the U.S. Federal Reserve of Treasury notes and mortgage-backed securities for the 14-year period from December 18, 2002 to December 18, 2016 in millions of dollars. Constructed by the author using data obtained from the FRED (Federal Reserve Economic Data) of the Federal Reserve Bank

of Saint Louis.

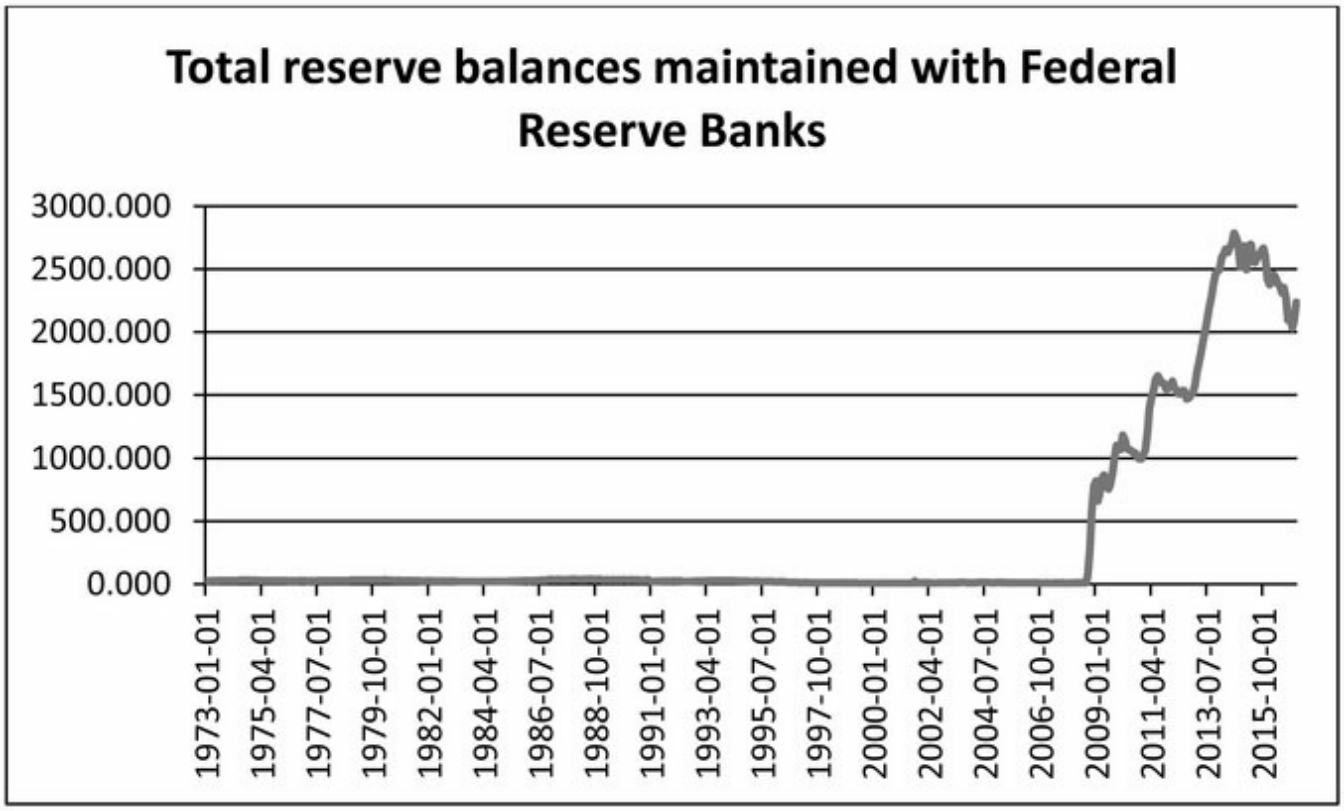

EXHIBIT 7.6 The evolution of reserves at U.S. commercial banks for the 42-year period from January 1973 to January 2015 in billions of dollars. Constructed by the author using data obtained from the FRED (Federal Reserve Economic Data) of the Federal Reserve Bank of Saint Louis.

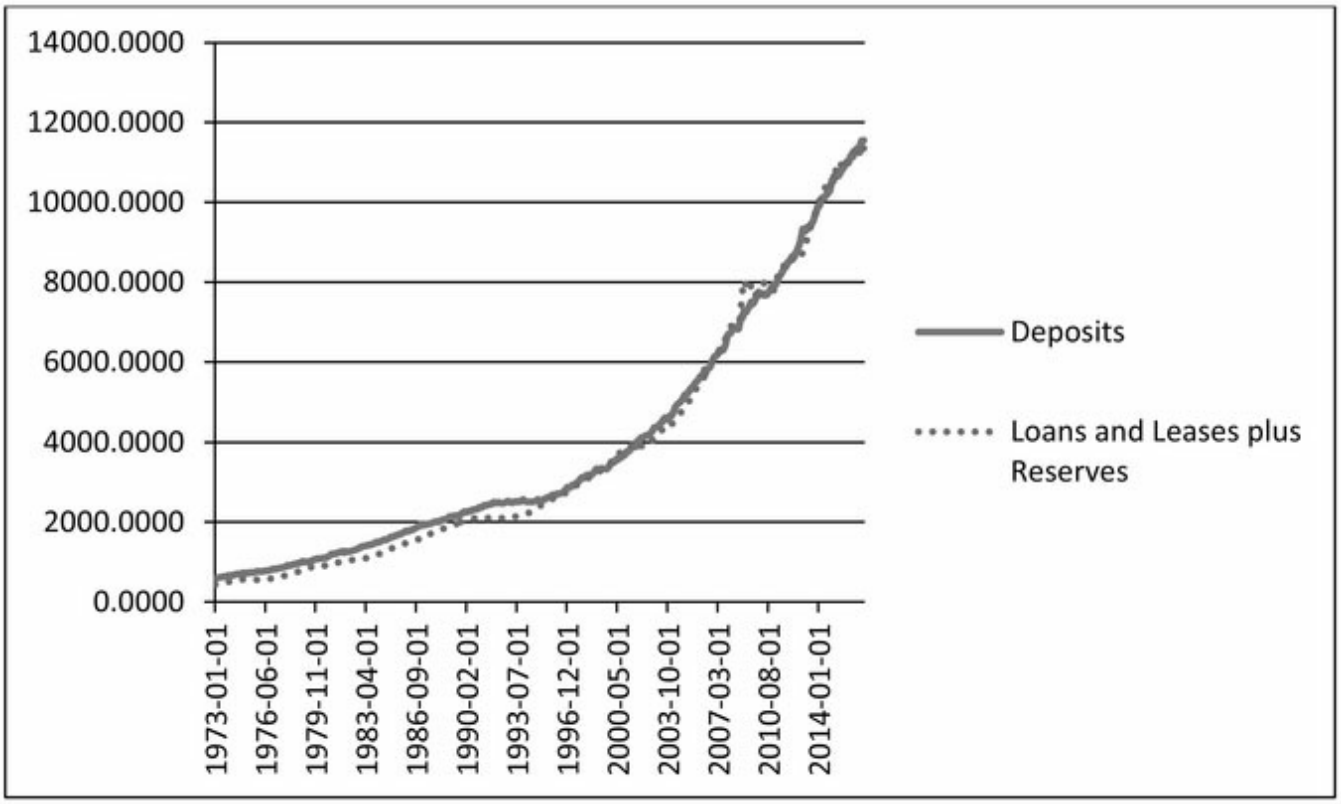

EXHIBIT 7.7 Comparison of the evolution of the total of deposits and of loans and leases plus reserves at U.S. commercial banks for the 44-year period from January 1, 1973 to January 1, 2017 in billions of dollars. Constructed by the author using data obtained from the FRED (Federal Reserve Economic Data) of the Federal Reserve Bank of Saint Louis.

These data illustrate the two realities of money created by commercial banks when they make loans (i.e. allocate new money) and money created by central banks when they buy assets from commercial banks (i.e. create reserves). When the central bank buys assets, there is no corresponding loan.

Let's try to "visualize" the working of money (i.e. its creation and allocation). In a large modern economy, outside of a small fraction in currencies, money resides in bank deposits. Deposits change continuously due to transfers of money from different entities, transfers implemented as transfers of deposits. At the same time, a large number of loans and leases are created and repaid daily. In Chapter 1 we saw that money serves three functions: money (1) facilitates transactions, (2) serves as a storage of value, and (3) serves as a unit of account. But, of course, if money is associated with loans it cannot, in aggregate, be a good storage of value: it does not make economic sense for an entity that stores value to earn low interest payments from an entity that borrows money.

Economists such as Silvio Gesell introduced the idea of money losing value in time. In Gesell's theoretical framework (actually put into practice as we saw in Chapter 3) the face value of money is depreciated regularly as notes in circulation must be stamped monthly for a fee – a sort of negative interest rate. How can credit money lose value? Earning little or no interest (or negative interest in Gesell's framework or when inflation is taken into consideration) makes money a bad investment. Holders of money might decide to use the currency only to perform transactions; the function of storage of value would be better served by putting the money elsewhere than deposits that are linked to loans and leases, for example financial assets such as money market funds or stocks.

7.4 Closing considerations on the role of commercial banks in allocating money

Today the preserve of commercial banks, credit is responsible for the creation and distribution of money. But are commercial banks really necessary for the creation and distribution of money? As we have seen, today money is created when an entity takes out a loan. Distributing new money only to those who take out loans is certainly one way to avoid issuing money for the profit of the issuing entity. In determining whom to lend money to, banks apply business rules, according credit to creditworthy individuals or entities. The latter are determined in function of a number of criteria which likely include a valuation of wealth and collateral and the feasibility of the borrower's project. In principle, such business rules should ensure the efficient allocation of credit money. But it is obviously not foolproof: mistakes can be made in evaluating creditworthiness; relationships, fads, or other can distort evaluations; collateral based on financial assets might prove to be self-referential.

As Augusto Graziani (1990) observed, money as credit calls for a system capable of ensuring that the generation of money functions properly. Today's answer is a banking system with commercial banks plus a central bank. Commercial banks ensure that no money is generated to the profit of someone without a corresponding debt. But the system needs trust. This trust has two sources: (1) prevailing social beliefs and values that underpin the present system and its rules and (2) regulations governing the banking system by national authorities such as the central banks and super-national authorities such as the Basel-based Bank of International Settlements, which is composed of 60 member central banks from nations responsible for some 95% of world GDP.

The banking system is in continuous evolution. The role of extending credit is being challenged by marketplace lending, as discussed in Chapter 5. A Morgan Stanley Research report (2015) predicted that marketplace lending could reach $290 billion in the U.S. by 2020. They based their forecast on an expected compound annual growth rate of 51% for the period 2014–2020. Advantages underpinning the growth were listed as fast response time, quicker loan approval times, faster funding, and lower borrowing rates. While the report found that "market-place lenders have focused on unsecured consumer credit, with roughly 80% of loans used to consolidate debt, and small business loans—with an estimated $100 billion in unmet demand solely in the U.S.," the report's authors suggested that U.S. marketplace lenders would likely move into other loan businesses including the $1.4 trillion student loan market, auto loans, and mortgages.

But there are additional factors that might challenge, in a more radical way, the role of commercial banks as creators and distributors of money through their role in issuing credit. These factors include the suppression of coins and banknotes, with payments being made exclusively through, for example, electronic payment systems or debit or credit cards. In 2015, the BoE kicked off a research program one of whose objectives was to consider the future of money. Among the motivations for the project were (1) low interest rates which were pushing central banks closer to the limits of conventional policy and (2) the rise of unconventional solutions to lending (marketplace lending) and money (digital currencies) with the potential to change how funds circulate through the economy. The question the BoE posed: What is the role of central banks if they were no longer needed to back state-issued currency? The future of money, BoE's chief economist and executive director for monetary analysis, research, and statistics Andrew Haldane commented, is about the most important challenge right now. Haldane linked the question of low to negative interest rates to cash: "There's an important cash dimension to [to ever lower interest rates] in that physical cash is one of the constraints that prevents interest rates going lower."[8]

In November 2016, the Indian government had a different set of motivations for withdrawing 86% of the country's currency overnight: to reduce corruption and fraud by reducing the circulation of anonymous, non-traceable money.

If it is now possible to envision an economy without currency in the form of coins and banknotes, can we envision a system where the allocation of new money is not linked to loans? Clearly the transfer of purchasing power between different agents as they exchange goods and services must be assured. Might commercial banks be replaced by "scoreboards" that record the transfers of purchasing power of each individual or entity? This would imply some sort of centralized clearinghouse and a supervisory entity to ensure that values are correctly transferred, or perhaps distributed ledgers implemented with blockchain technology as presently used by digital coin builders. That scoreboard, clearinghouse-cum-supervisory entity, or distributed ledger where money is purely a number in a computer system might be managed by the government, a central bank, or some sort of banking system.

Ben Broadbent (2016), deputy governor for monetary policy at the BoE, discussed a possible adoption of blockchain technologies (he prefers the term "decentralised virtual clearinghouse and asset register") in the implementation of central bank digital currencies (see Chapter 3 for a discussion of the technological issues associated with central bank digital currencies with respect to private digital currencies). But central bank digital currencies raise important economic questions. Broadbent envisaged the possibility of allowing non-banks direct access to accounts

with the BoE:

> One might go further, giving access to non-financial firms, or perhaps even individual households. In the limit, a distributed ledger might mean that we could all of us hold such balances . . . If so, our accounts would no longer be a claim on commercial banks but, like banknotes, the liability of the central bank.

Discussing the nature of these accounts, Broadbent asked if central bank accounts should entail services such as the payment of interest or simply be "e-cash" used for retail transfers without interest payments.

In function of the nature of the accounts, the problem of allocation of central bank digital money would be critical. Today new money is created by commercial banks in extending loans; banknotes – though only a small portion of the money in circulation – are distributed to clients debiting their bank accounts; new private digital currencies (Bitcoins are estimated to be worth $138 billion at the end of November 2017) are presently created and distributed as the remuneration for miners who perform the shared open record-keeping service for the currency (see Chapter 3). Should central banks play a key role in the creation and distribution of digital currencies, how might the present rules governing the creation and distribution of money change? Will the brave new world of robotization create new social or economic pressures, calling for new rules? Might deposits be made permanent in aggregate, with an equal amount of money allocated periodically to all citizens? This system, however, would make investments difficult as it would lack the capability of creating upfront large sums of money and would, likely, be strongly opposed.

Notes

1 Melvin J. Lerner, a social psychologist associated with the University of Waterloo, conducted seminal research on the notion of justice. In his article "The Justice Motive: 'Equity' and 'Parity' among Children," Lerner (1974) reported on three experiments with children's use of equity and parity forms of justice in determining the allocation of rewards. The first involved 60 kindergartners, defined as a team. According to Lerner, the kindergartners exhibited little evidence of self-interest and closely followed the rule of parity in distributing the rewards. Another experiment involved 90 5th graders – defined as coworkers rather than team members – who exhibited the use of the justice of equity as well as parity. A third experiment with 80 1st graders tested more directly the importance of being defined as coworkers (nonunit) or team members (unit) as a determinant of the equity or parity forms of justice. Results suggest that children are highly motivated to follow rules of justice. More recent studies have underlined children's notions of justice as regards starting-point allocations and rewards.

2 As we have seen, central banks have no relationships with private entities, so the only possible exception to giving or lending banknotes to commercial banks is giving or lending banknotes to the government. But, as we have noted, this is not allowed in major Western countries including the U.K. and the U.S. and in the euro zone.

3 See https://www.boj.or.jp/en/announcements/education/oshiete/outline/a23.htm/

4 As already mentioned, transfers, for example for payment of goods or services, do not create money but only transfer existing money.

5 This section draws on McLeay (2014a, 2014b).

6 In addition to aggressive bond purchases, the Bank of Japan (BOJ) announced a $58 billion-a-year stock purchase program in 2016, making the Bank one of the country's biggest stock market investors. Bloomberg (2016) estimated that by the end of 2017, the BOJ could become the largest shareholder of 5[illegible] companies in the country's Nikkei 225.

7 Open market operations consist in a central bank buying or selling Treasury debt to commercial banks thereby increasing or reducing commercial banks' reserves.

8 Hamilton and Charlton (2016).

References

Bloomberg. 2016. "BlackRock Sides with BOJ in Debate over Tokyo Whale's ETF Buying." (September 15): www.bloomberg.com/news/articles/2016-09-14/blackrock-sides-with-boj-in-debate-over-tokyo-whale-s-etf-buying.

Broadbent, Ben. 2016. "Central Banks and Digital Currencies." Speech given at the London School of Economics. (March 2): www.bankofengland.co.uk/publications/Pages/speeches/2016/886.aspx.

Ferguson, Niall. 2008. *The Ascent of Money: A Financial History of the World*. New York: The Penguin Press.

Friedman, Milton. 1969. "The Optimum Quantity of Money" in *The Optimum Quantity of Money and Other Essays*. London: Aldine.

Graziani, Augusto. 1990. "The Theory of the Monetary Circuit." *Economies et Societies*, vol. 24, no. 6:7–36.

Grazzini, Enrico. 2016. "Why Fiscal Money Is Better than Helicopter Money and QEP at Beating Deflation and Austerity." *Social Europe*. (November 8): www.socialeurope.eu/fiscal-money-better-helicopter-money-qep-beating-deflation-austerity.

Hamilton, Scott and Emma Charlton. 2016. "Haldane's Brave New World at Bank of England Probes Fate of Cash." *Bloomberg*. (February 2).

Jácome, Luis I., Marcela Matamoros-Indorf, Mrinalini Sharma, and Simon Townsend. 2012. "Central Bank Credit to the Government: What Can We Learn from International Practices?" IMF Working Paper. (January).

Keen, Steve. 2011. *Debunking Economics: The Naked Emperor Dethroned?* Revised and expanded ed. London: Zed Books.

Lerner, Melvin J. 1974. "The Justice Motive: 'Equity' and 'Parity' among Children." *Journal of Personality and Social Psychology*, vol. 29, no. 4 (April):539–550.

Martin, Felix. 2013. *Money: The Unauthorized Biography – From Coinage to Cryptocurrencies*. London: The Bodley Head/Random House Group.

McLeay, Michael, Amar Radia, and Ryland Thomas. 2014a. "Money in the Modern Economy: An Introduction." *Quarterly Bulletin* (Q1). The Monetary Analysis Directorate of the Bank of England.

McLeay, Michael, Amar Radia, and Ryland Thomas. 2014b. "Money Creation in the Modern Economy." *Quarterly Bulletin* (Q1). The Monetary Analysis Directorate of the Bank of England.

Morgan Stanley Research. 2015. "Can P2P Lending Reinvent Banking?" Morgan Stanley Blue Papers. (June 17): www.morganstanley.com/ideas/p2p-

marketplace-lending.
Mosler, Warren. 2010. *The Seven Deadly Innocent Frauds of Economic Policy*. Christiansted, USVI: Valance Co.
Ryan-Collins, Josh, Tony Greenham, Richard Werner, and Andrew Jackson. 2011. *Where Does Money Come From?* London: New Economics Foundation.
Werner, Richard A. 2016. "A Lost Century in Economics: Three Theories of Banking and the Conclusive Evidence." *International Review of Financial Analysis* 46 (2016):361–379.
Wray, L. Randall. 2015 [2012]. *Modern Money Theory*. 2nd ed. [1st ed.]. London: Palgrave Macmillan.

8

Money and the Economy

In previous chapters we discussed the nature and dynamics of money. Here we discuss the impact of money on the real economy. By necessity we depart from mainstream economic theory in that the latter is an equilibrium theory that gives little or no consideration to the interaction between the dynamics of money and the real economy. However, it seems difficult not to recognize that the dynamics of money has a real impact on the economy.

Heterodox economists have begun to consider the problem of the integration of the theory of money with macroeconomics in a non-equilibrium setting. One such attempt is proposed in the stock-flow consistent models of economies with a banking system. We discuss this and also argue that the proper integration of the theory of money and macroeconomics can be achieved only in the context of the theory of complexity.

This chapter begins with a discussion of the difficulties encountered when attempting to integrate money and a banking system into classical macroeconomics. Circuitism will be discussed as it gives a rigorous theoretical framework to the notion that money is a tool borrowed to perform transactions. We discuss a number of questions that are not explained within the classical framework. Among these are the puzzles of the reduced velocity of circulation of money and of missing inflation. The theory of financial and economic crises based on Minsky's analysis will also be discussed. We suggest that a proper understanding of the impact of money on economies requires a rethinking of macroeconomics, borrowing concepts from the theory of complexity because in modern complex economies the distinction between real and nominal quantities is somewhat blurred.

8.1 Money and classical economic theory

Why is the integration of money with mainstream macroeconomics difficult? Two factors explain the difficulty:

- Mainstream economics is an equilibrium theory while the dynamics of money might lead to instabilities;
- Mainstream economics fails to recognize the complexity of modern economies.

Mainstream economics is embodied in Dynamic Stochastic General Equilibrium (DSGE) models. (SeeBox 4.1 for an explanation of General Equilibrium Theories.) Following the so-called "Lucas critique," DSGE models model economic dynamics based on forecasts made by the economic agents populating the economy. Robert Lucas (1976), an American economist and recipient of the 1995 Nobel Memorial Prize in Economic Sciences, observed that in modelling the effects of policy changes we have to consider that agents will anticipate the changes, thereby rendering them ineffective. Lucas advocated giving macroeconomics a microeconomics foundation, that is, creating macroeconomic models based on the behavior of agents. Lucas was influential in the formulation of neoclassical economics which is based on three main concepts: agent optimality, equilibrium, and rational expectations.

According to neoclassical economics, real economies are populated by agents who make forecasts that are in perfect agreement with the actual development of economic variables. Agents are said to have rational expectations, a concept introduced by the American economist John Muth (1961). Based on their rational expectations, agents optimize their economic decisions.

Rational expectations imply equilibrium in the sense that the collective decisions of agents must produce the forecasted evolution of variables, for if not, forecasts would be invalidated. This is a logical requirement: if an intelligent processor of information receives a forecast about variables that are controlled by its own behavior then, to be credible, the forecast must be in agreement with the intelligent processor's decisions. Real economic agents, however, do not optimize based on rational expectations. In fact, real economic agents make decisions based on information from the past and make forecasting errors. Eric Beinhocker (2007), executive director of the Institute for New Economic Thinking's research program at the University of Oxford (INET@Oxford),[1] illustrates the dilemma of a rational agent who goes grocery shopping:

> You have well-defined preferences for tomatoes compared with everything else you could possibly buy in the

> world, including bread, milk, and a vacation in Spain. Furthermore, you have well-defined preferences for everything you could possibly buy at any point in the future, and since the future is uncertain, you have assigned probabilities to those potential purchases. For example, I believe that there is a 23% chance that in two years the shelf in my kitchen will come loose and I will need to pay $1.20 to buy some bolts to fix it. The discounted present value of that $1.20 is about $1.00, multiplied by a 23% probability, equals an expected value of 23 cents for possible future repairs, which I must trade off with my potential purchase of tomatoes today, along with all of my other potential purchases in my lifetime. . . . [To make your decisions,] you know exactly what your budget is for spending on tomatoes. To calculate this budget, you must have fully formed expectations of your future earnings over your entire lifetime and have optimized your current budget on the basis of that knowledge. In other words, you might hold back on those tomatoes because you know that the money spent on them could be better spent in your retirement. Of course, this assumes that your future earnings will be invested in a perfectly hedged portfolio of financial assets and that you take into account actuarial calculations on the probability that you will live until retirement at age 65, as well as your expectations of future interest rates, inflation, and the yen-to-dollar exchange rate. While standing there, staring at those nice, red tomatoes, you then feed all this information into your mind and perform a cunning and incredibly complex optimization calculation that trades off all these factors, and you come up with the perfectly optimal answer – to buy or not to buy!
>
> *(p. 116)*

This description might look like a caricature, but it is exactly what is implied by a rational expectations model. A rational-expectation agent cannot simply buy tomatoes because they look appetizing and the price is good but has to weigh the purchase of tomatoes against any other possible event. Economists such as Von Mises and Hayek criticized Marxist-type command economies on the basis, among others, that the task of gathering and processing all the information needed to plan an economy is excessively difficult. But the quintessential free-market models such as DSGE assign each individual agent a task of similar daunting complexity.

As for equilibrium, the frequency and magnitude of recent financial crises have seriously questioned this hypothesis for both markets and economies. As mentioned earlier, Romer (2013) counted six distinct shocks in U.S. markets during the past 30 years or so that have posed important macroeconomic risks. Hyman Minsky (1986) argued that the emergence of a crisis is facilitated by excess money which might create disequilibria. As will be discussed later in this chapter, money as credit creates disequilibria essentially because it creates local loops of positive feedback. This might occur as new money creates demand in excess of available assets, thereby causing a rise in asset prices which, in turn, allows the creation of ever more money. Money can therefore produce bubbles as asset prices grow at rates significantly higher than the growth rate of the real economy; it might additionally create rapid nominal growth in subsectors of the economy. As money is generated through credit, excess money might create a situation where debts cannot be repaid, thereby triggering a crisis.

The existence of disequilibria and of financial and/or economic crises is not universally accepted. Some prominent economists argue that crises are unpredictable

random events. In the aftermath of the 2007–2008 financial market crisis, Lucas (2009) wrote, “One thing we are not going to have, now or ever, is a set of models that forecast sudden falls in the value of financial assets.” Lucas was not alone in his evaluation. Even when a crisis does occur such as the 2008 subprime mortgage crisis, neoclassical economists can be quite glib about it. Robert Barro (2009), professor of economics at Harvard University, remarked during a roundtable discussion organized by *The Economist*, “Economies have natural tendencies to recover from recessions, and such a recovery is the most likely outcome for the American economy going into 2010.” The problem for classic economic theory is that situations of disequilibrium present a major departure from classical macro-modelling: situations of disequilibrium would be required by the introduction of money as credit.

8.2 The Theory of the Circuit of Money

In Chapter 1 we defended the need for a theory of money based on the fact that money has a dynamics of its own and that that dynamics impacts the real economy. Here is how. Money is created and allocated with rules independent from the real economy. Money allocates purchasing power and therefore enables and transmits major economic decisions. From the point of view of consumers, money affects consumption decisions; from the point of view of capitalists, money affects investment and production decisions. A discussion of the impact of money on the real economy calls for an understanding of the dynamic process of the creation, distribution, and destruction of money. This, in turn, calls for a discussion of the Theory of the Circuit of Money, or Circuitism.

Circuitism, in fact, presents a neat and rigorous theoretical framework for understanding the dynamics of money as credit, with delays and feedbacks. As already observed, money as credit is not permanent but is borrowed to execute transactions. Circuitism gives a rigorous theoretical foundation to this idea. Though Circuitism has had to change to adapt to new empirical realities, it still offers a conceptual benchmark to the dynamic theory of money. Let's start by looking at classical Circuitism; successively we will explore the reasons that called into question its original formulation and its subsequent reformulation.

8.2.1 Classical Circuitism

Circuitism is a heterodox (i.e. non-classical) macroeconomic theory that embraces a theory of money. Indebted to ideas from Marx, Wicksell, Keynes, and Minsky, Circuitism as a theory of money was proposed simultaneously in France by Alain Parguez and in Italy by Augusto Graziani (Realfonzo [2012] surveys the literature on the Italian Circuitist approach). More recently new versions of Circuitism have been proposed, including by the Canadian economists Lavoie, Seccareccia, and Rochon.

Let's begin with a discussion of the formulation of Circuitism as presented in Graziani (1990). Though no longer supported by empirical data, we will discuss Graziani's formulation for didactical reasons: it provides, as mentioned earlier, a clear exposition of the notion of a monetary circuit that is the benchmark against which dynamic theories of money must be evaluated.

Circuitism is based on a number of macroeconomic conceptual pillars. Consider a closed economy. Circuitism assumes that in economies there are workers and capitalists; this is the first macroeconomic conceptual pillar of Circuitism. Workers

earn wages; capitalists earn profits. In modern economies, capitalists are no longer a well-defined entity. In fact, a distinction now has to be made between those who manage firms (the decision-makers) and those who earn the profits, typically as shareholders. Decision-making might, in turn, be split between the managers of a firm and the professional asset managers who are the direct owners of firms. This situation might create the so-called "agency problem" as a firm's managers might make decisions that benefits themselves and not the owners. However, from the point of view of Circuitism as originally exposed by Graziani, the complex structure of the ownership (including shareholders) and the management of modern firms does not need to be taken into consideration, at least as a first approximation. The fundamental point is that there is a separation between workers who earn wages and capitalists who earn profits, where we might group together owners and managers and call them capitalists. We can do so because in modern economies a large fraction of the ownership of firms is concentrated in the hands of a small group of asset managers[2] and/or of individuals or families. Investment funds, be they pension funds or mutual or trust funds, are typically profit-seeking capitalists though profits are shared by a large number of people.

The second macroeconomic conceptual pillar of Circuitism assumes that money is bank credit, endogenously created by the banking system, supplied on demand (and under certain guarantees) to firms that request it. This is a major departure from classical theory: in classical macroeconomics, firms are limited by their savings in making new investments. Not according to Circuitism. Circuitism maintains that having a feasible business project, firms are not constrained by lack of money; they can borrow all the necessary funds. This is a consequence of the fact that credit money is endogenously generated – now a widely accepted concept.

The notion that money can be created *ex nihilo* on demand to respond to investment needs might seem like magic. Actually it is the contrary. In theory, the creation of money on demand allows the coordination of productive activities when resources are available but the money to coordinate them is missing. That is to say, monetary savings might not balance the availability of human and physical resources, in which case the money needed to coordinate resources towards productive ends is lacking. The creation of money as credit solves the problem. Nevertheless, a lack of money for investment might occur if, for example, a central bank were to control the quantity of money through reserve requirements, limiting banks' ability to provide sufficient credit.

Circuitism then observes that production and sales take time to develop, leading to its third macroeconomic conceptual pillar. According to Circuitism, the circuit of money functions in the following way. First, firms borrow money to finance

production. This money is created *ex nihilo* by the banking system. Capitalists use the money to pay wages. At the end of the production cycle, which might be very different in function of the firm or the industry, a fraction of manufactured goods are sold to the workers. Workers purchase goods with their wages; the money returns to capitalists. Capitalists can then repay their loans, thereby destroying the borrowed money.

Graziani adds a distinction between initial and final finance. Initial finance is effectively the money borrowed to finance production. Final finance is the process through which money comes back to firms as workers spend their wages buying goods and services. At the end of the cycle, workers might decide to save some portion of their wages, which they will put in their bank account, or buy debt (e.g. bonds) or corporate shares. If workers buy new debt or new shares, the money goes back to firms and capitalists who use the money to repay their loans. If, on the other hand, workers decide to keep some money in their bank accounts, then this money does not return to firms and firms cannot repay their debt. In the language of Circuitism, there is a leak in the circuit.

Classical Circuitism considers money as something borrowed from banks by capitalists or firms which is then used to coordinate production and pay wages, and finally returned to banks after products are sold. In real production processes there are, of course, intermediate manufacturing steps before reaching consumers and therefore there are additional local circuits. Consider that the production of a car includes the assembly of, among other things, tires and the body of the car. Tire manufacturers pay wages and sell tires to car manufacturers. The tire manufacturers are paid by the car manufacturers. Therefore the income generated by the sale of cars goes to pay wages but also to pay suppliers which, in turn, pay wages.

There are a number of issues in Circuitism as exposed above. First, workers car pay back at most what they have received as wages. This leads to a paradox: the apparent impossibility (in the aggregate) of creating monetary profits for the capitalists or even of the capitalists paying interest on loans. To see this, let's use a well-known notation, initially used by Marx: capitalists borrow a sum M, to pay wages, their firms produce an output C, the output C is sold to workers who spend M'. In aggregate, the profit P of the capitalists is $P = M' - M$. The paradox is given by the fact that workers cannot spend more than they receive as wages and it is therefore not possible that $M' > M$, that is, it is not possible that $P > 0$.

Box 8.1 Marxism and Circuitism

The division, and ultimately the opposition, of capitalists and workers in Circuitism might seem similar to the Marxist class war. Note, however, that in the Marxist system class war, which had its philosophical roots in Hegel's dialectics, was intended as a serious conflict whose evolution would follow a sort of dialectic determinism, leading to the final synthesis of a new social order where class divisions would be eliminated. This component is absent from Circuitism. Though an author like Graziani might have been influenced by Marxism, Circuitism in itself is a purely economic theory.

Empirically, in modern economies there are wages and profits. We might also consider undeniable the fact that the distribution of profits is highly skewed, that is, it is concentrated in the hands of a small fraction of the population. Using data from 20 countries going as far back as the 18th century, Thomas Piketty (2014) analyzed the distribution of wealth and profits. His conclusion: the distribution of wealth is highly concentrated. The exact parameters of this distribution have been contested but Piketty's conclusion is undeniable. Setting aside those who believe that modern capitalist economies are more egalitarian than they actually are, the main political debate is whether high concentration of wealth (and power) is to be considered positive, in that it proves that merit is rewarded, or bad, in that the cards are stacked against those on the lower rungs of the ladder.

The object, however, of our discussion is the theory of money and its possible impact on the economy. Circuitism offers a basic theoretical framework for analyzing economies where economic decision-making is concentrated in the hands of a small fraction of the population, that part of the population that earns profits. Some of the original assumptions of Circuitism have been reformulated to reflect changes in borrowing patterns, that is, the fact that today it is the wage-earners that are the principal borrowers.

The question of monetary profit has been widely discussed in the literature. Charlotte Bruun and Carsten Heyn-Johnsen (2009) note that the paradox of monetary profits has been a recurrent theme in macroeconomics ever since it was first formulated by Marx. The argument is that capitalists in aggregate cannot get more money back from workers than what the workers in aggregate are paid in wages. Marx was not able to solve this problem and neither was Keynes, who struggled with it in formulating his General Theory. Bruun and Heyn-Johnsen write that a consequential logical conclusion to Keynes' treatment of the problem renders indeterminate his concept of aggregate income because it is based on "imaginary

magnitudes."

The conclusion that in aggregate there is no monetary profit seems unescapable. If all money created is used to pay wages and must be returned and destroyed there is no way to accumulate money as profit. This does not mean that firms cannot show an economic profit. Suppose that, in aggregate, capitalists borrow money and buy the extra goods and services produced by their own firms that wage earners cannot afford with their wages. Firms will make a profit via these sales and will pay dividends to capitalists. Capitalists get their money back and can repay loans. In the end, firms will have shown economic profits, capitalists will have received cash dividends and acquired goods and services. Capitalists earn a profit in aggregate; this profit is not money but material wealth embodied in objects such as houses, cars, and yachts. Firms exhibit profits but immediately pay dividends to the capitalists. The point is that the production and sales processes do not create new money: they create goods and services that can be purchased with borrowed money. If households do not borrow, and therefore do not create money, the profit of capitalists, in aggregate, cannot be money but only goods and services.

The Circuitist notion that capitalists borrow from banks to finance production is a conceptual idealization that does not reflect the present situation. As observed by Adair Turner (2013), today firms are not the main borrowers, and therefore not the main generators of money. The largest part of debt today is taken on by households, that is, by the general public. It is thus the general public that originates most of the newly created money that finances production and in the end allows capitalists to make a financial profit.

A second issue with Circuitism is mathematical modelling. There is no full-fledged mathematical theory of Circuitism as originally proposed. More recently, attempts have been made to create Circuitist models using System Dynamics[3] within the general framework of stock-flow consistency proposed by Godley and Lavoie (2007). But modelling the circuit of money is made difficult by the problems associated with monetary profits and the payment of interest.

8.2.2 The changing landscape of borrowing

Circuitism as we have discussed might have better reflected the reality of bank lending in the 1950s and 1960s [4] when it was initially conceived but certainly it does not reflect the current distribution of bank lending. In fact, in the original version of Circuitism, the demand for bank lending originates from firms while workers are basically not indebted. But this is not empirically the case in modern capitalistic

economies as discussed in Turner (2013). To illustrate this, let's look at bank lending in the U.S. The situation in the U.S is similar to that of England and of most members of the European Union. Total bank lending in the U.S. is illustrated in Exhibit 8.1. We can see that at the time of the crisis in 2007, new lending stopped and started again in 2011 though at a lower rate.

Exhibit 8.2 illustrates lending to industrial and commercial firms.

Exhibit 8.2 shows significant fluctuations with peaks before the crises of 2000 and 2007. Exhibit 8.3 shows the ratio of industrial and commercial loans to all loans. As can be seen in this exhibit, there was a marked trend downward, from a high of 42% in 1947 to a low of 17% in 2010. The ratio was up in 2017, to *c* .23%.

Let's now look at loans in the real estate sector. Exhibit 8.4 shows an acceleration in the growth of real estate loans, reaching a peak just before the crisis of 2007, after which it declined though by 2017 real estate loans were again at the same level as in 2007.

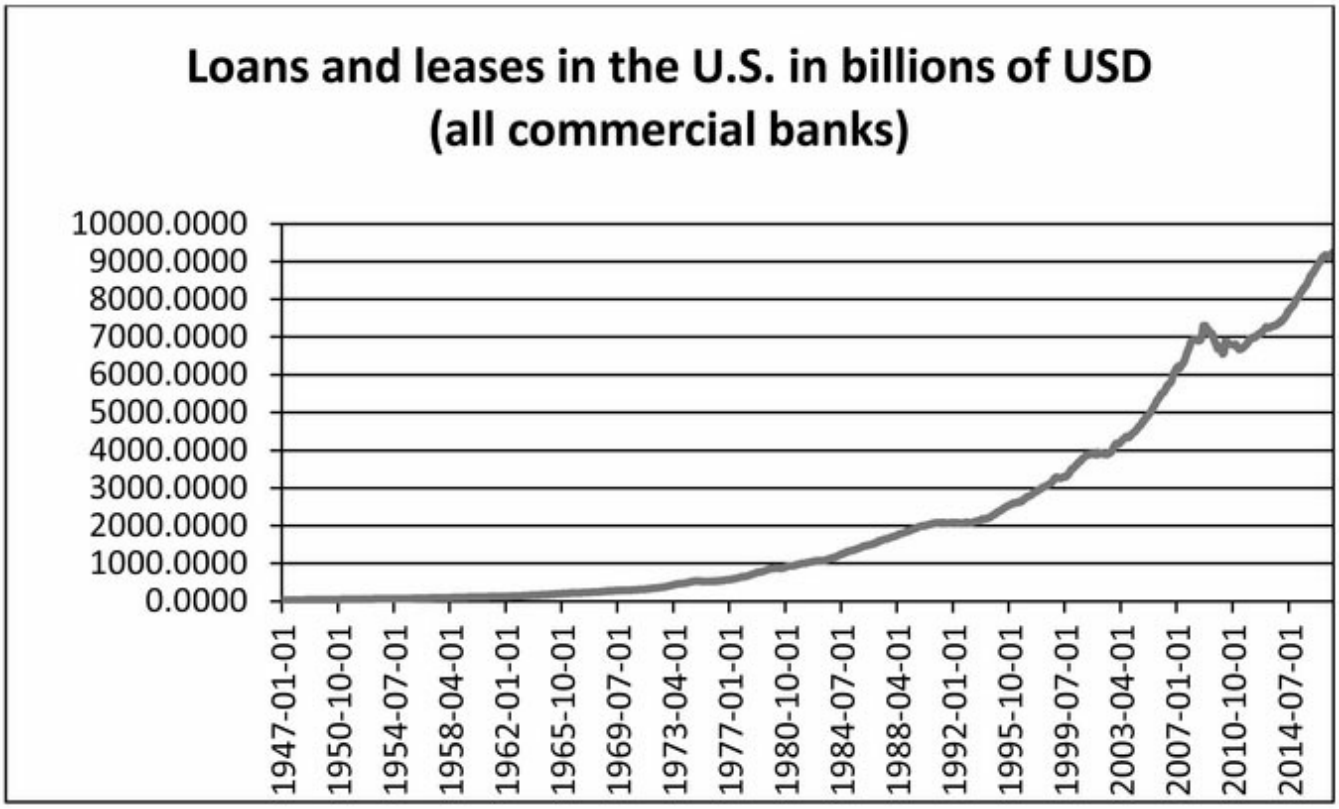

EXHIBIT 8.1 All loans and leases in the U.S. in billions of USD for the 70.5-year period from January 1947 to June 2017. Constructed by the author using data obtained from the FRED (Federal Reserve Economic Data) of the Federal Reserve Bank of Saint Louis.

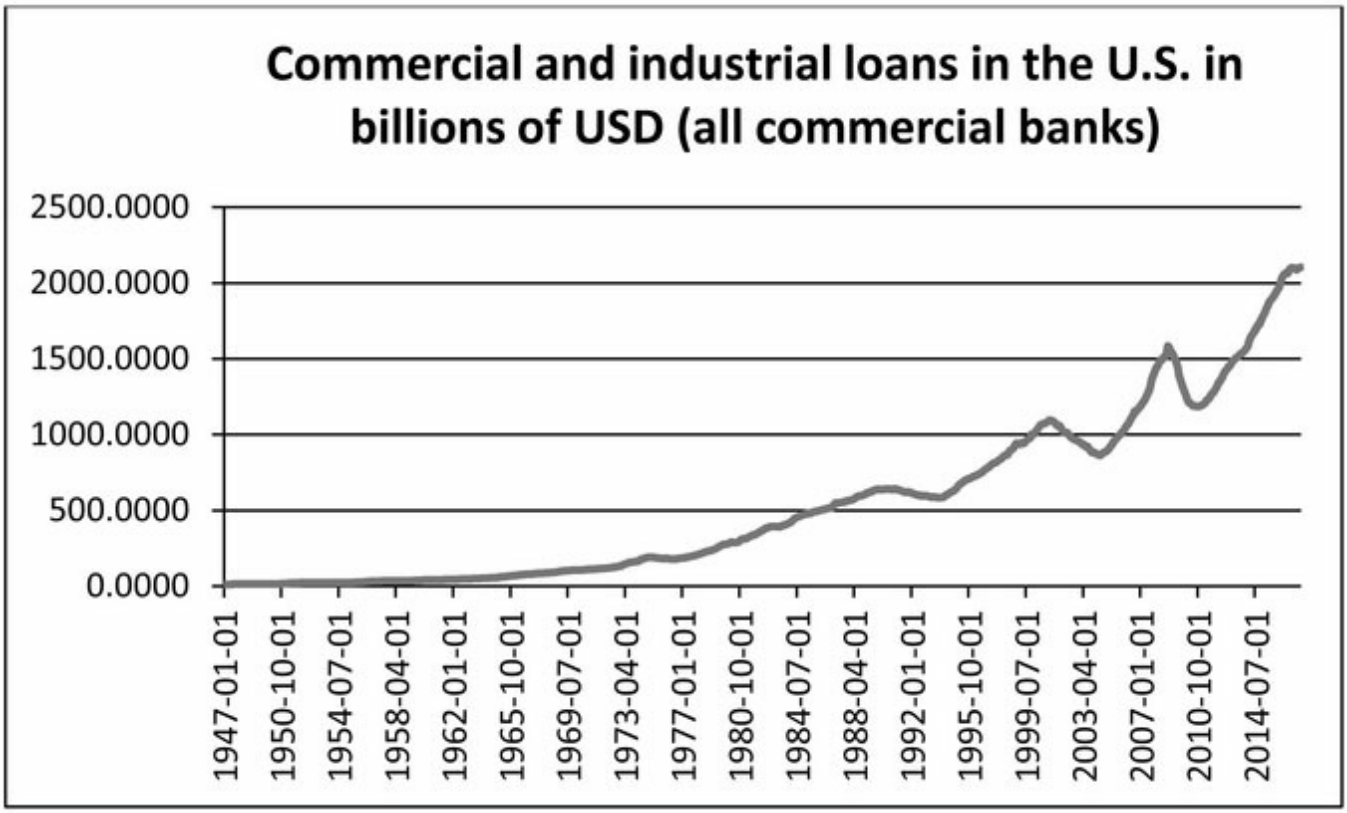

EXHIBIT 8.2 All commercial and industrial loans in the U.S. in billions of USD for the 70.5-year period from January 1947 to June 2017. Constructed by the author using data obtained from the FRED (Federal Reserve Economic Data) of the Federal Reserve Bank of Saint Louis.

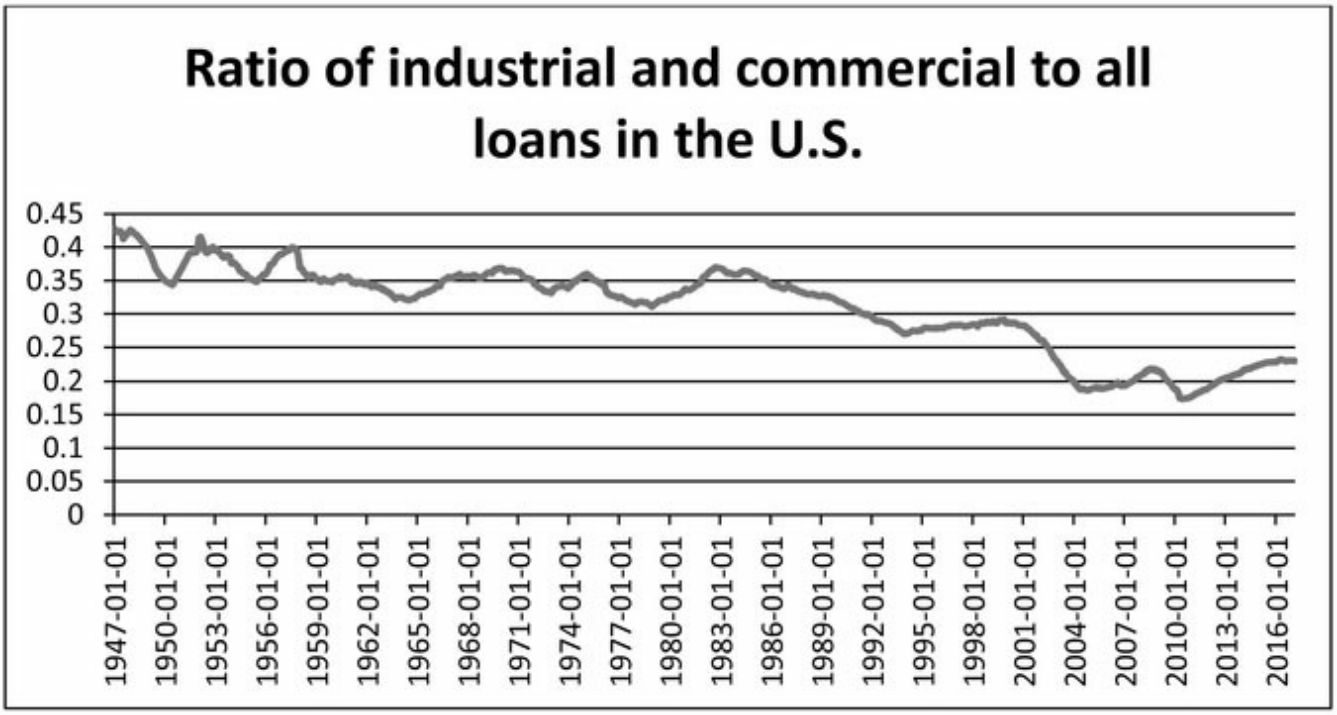

EXHIBIT 8.3 Ratio of industrial and commercial loans to all loans in the U.S. in the 70.5-year period from January 1947 to June 2017. Constructed by the author using data obtained from the FRED (Federal Reserve Economic Data) of the Federal Reserve Bank of Saint Louis.

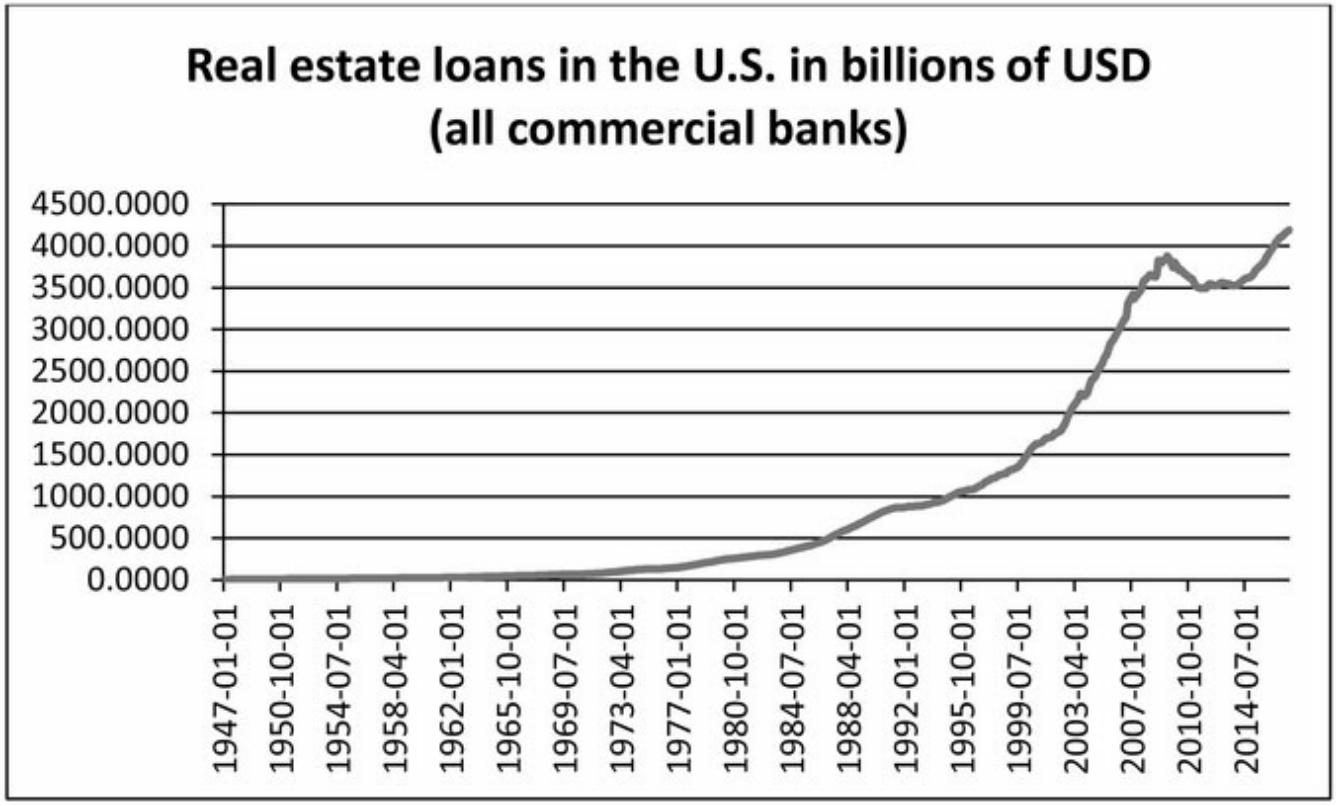

EXHIBIT 8.4 All commercial and industrial loans in the U.S. in billions of USD for the 70.5-year period from January 1947 to June 2017. Constructed by the author using data obtained from the FRED (Federal Reserve Economic Data) of the Federal Reserve Bank of Saint Louis.

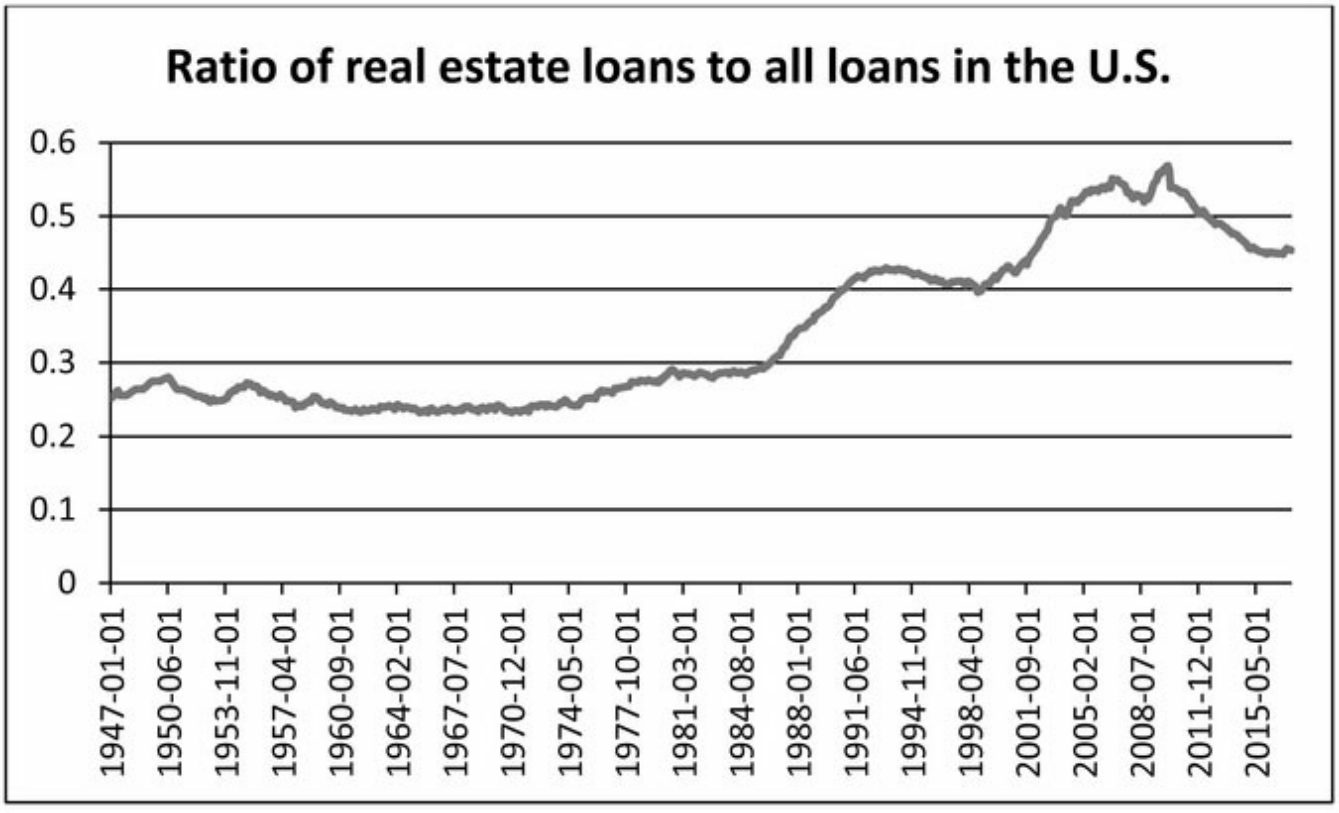

EXHIBIT 8.5 The ratio of real estate loans to all loans in the U.S. for the 70.5-year period from January 1947 to June 2017. Constructed by the author using data obtained from the FRED (Federal Reserve Economic Data) of

the Federal Reserve Bank of Saint Louis.

Exhibit 8.5 illustrates the ratio of real estate loans to all loans. The ratio was 25% in 1947, then grew to 55% in 2007 before declining to 45% in June 2017.

Exhibits 8.1–8.5 illustrate that today industrial and commercial loans are not the primary source of demand for credit. Real estate lending has a bigger share of total lending than industrial and commercial. In June 2017, at $4.19 trillion, real estate lending is close to 50% of total borrowing, while at $2.1 trillion industrial borrowing is slightly above 20%. In addition to real estate loans in the U.S., there were $1.45 trillion outstanding in student loans and $1.13 trillion in motor vehicle loans in June 2017. In figures, this translates into a total outstanding U.S. corporate debt of $2.1 trillion in June 2017 compared to $6.6 trillion in real estate loans, student loans, and auto loans combined. That is to say, U.S. consumer debt is more than three times U.S. corporate debt. The last three decades have seen a progressive reduction of corporate borrowing as corporate profits have increased (see Chapter 1). Meanwhile, the general public first reduced their savings, from a high of 13.9% of disposable income in mid-year 1975 to a near low of 3.8% of disposable income in 2017 (see https://fred.stlouisfed.org/series/PSAVERT) – all the while increasing their borrowing. The general public is now a net borrower, firms net lenders.

Clearly the money circuit as described by the Circuitists, which had the merit of clearly describing the process of industrial financing with credit, needs to be revised. In addition to changes in the profile of the borrowers, we have to take into consideration the growing financialization of economic activity. Financialization is an ill-defined term: roughly speaking, it involves recourse to markets using engineered contracts as opposed to financing from banks. Financialization has a number of consequences as firms might store their profits using near-money or other liquid assets.

8.2.3 Circuitism revisited

Several changes call for a reformulation of Circuitism. First, as discussed earlier, firms, in aggregate, are no longer the principal borrower in most developed economies; households are now the principal borrower. Second, central banks' use of quantitative easing (QE) has created huge sums of money not linked to loans. Third, shadow banking has played a role in the development of liquid financial assets used as storage of value. Lastly, financial markets are playing a larger role in financing firms. Seccareccia (2012), Passarella (2014), and Sawyer (2013) coun

among those who have made an attempt to reformulate Circuitism, adapting it to the new empirical framework. In the classical Circuitist framework, there are three principal actors: firms, households, and banks; in the new Circuitist framework there are four principal actors; firms, households, banks, and financial markets. In essence, the reformulation of Circuitism consists primarily in the inclusion of financial markets and the reversal of the classical cycle where firms are the borrowers and households the savers, into a more complex system where firms are the savers and financial investors and households the debtors.

Let's look a bit closer at the current situation. If we take a snapshot of money stock in the U.S. in June 2017 we see that a large fraction of bank deposits is associated with household loans. In simple terms, money is created by people buying homes, motor vehicles, education, and other consumer items. Loans that have created this money are continuously being repaid and new loans extended. The new money thus created adds to the aggregated purchasing power. It enters into circulation and extends the purchasing power of households beyond the limits of their wages. Firms can therefore make monetary profits as they can sell more than the total wages they pay. Profits are then distributed to capitalists as dividends. This is a departure from classical Circuitist models where households spend only their wages and do not borrow.

Firms sell production and services to households and capitalists alike. Sales are financed by both wages and profits. Debts must be repaid, but as long as the total amount of household debt is growing it allows households to consume in excess of their wages and capitalists to increase their purchasing power. Corporate profits might be distributed or kept in the firm to finance operations and/or investments. Retained profits will typically be invested in a variety of near-money or other liquid investments. Profits distributed as dividends can be used to finance consumption or be reinvested in financial markets.

The simple circuit of money of classical Circuitism has now become a complex web of flows of money originated principally by household borrowing. As argued in Werner's Quantity Theory of Credit, the new money is split between money that finances consumption and money that is invested in financial markets. However, the central idea of Circuitism remains valid: money is a tool borrowed to perform transactions after which it must be returned and destroyed. Michell (2016) maintains that the theoretical framework of initial and final finance remains a useful conceptual pillar even in the new empirical setting.

The current circuit of money is based on a continuously growing amount of borrowing. Is this process sustainable? Will households be able to repay an increasing load of debt given stagnant wages? Cauwels and Sornette's (2012) paper

"The Illusion of the Perpetual Money Machine" argues that the process of increasing debt without increasing real gains is not sustainable. The key lesson of Circuitism is that money creation is a process that starts with creation of money through loans and ends with destruction of money as the loans are repaid. With stagnant wages it is difficult to see how consumers can repay an increasing amount of debt. Circuitism shows that the circuit must close.

Reformulated Circuitism has its critics. In particular, Lysandrou (2014) objects that the aggregation in main blocks – firms, households, banks, and now financial markets – is not methodologically correct. In addition, Lysandrou makes a number of empirical observations that tend to prove that it is not true that firms are lenders and not borrowers, observing that firms issue bonds – a substantial part of their borrowing. Lysandrou argues that the idea that bank-based credit relations are the dominant component of all money is not legitimate and proposes a different framework based on money as a commodity. He maintains that the position of domination is occupied by the impersonal commodity exchange relation, not the personal credit relation. Lysandrou writes:

> corporate bonds and corporate bank loans are simply alternative ways in which corporations raise money from creditors. . . . bonds and bank loans represent two very different types of debt: bonds notionally come under the commodity principle insofar as they signify the compression of the debt relation into a tradable entity while bank loans do not come under this principle because they signify a credit relation between fixed counterparties.
>
> *(p. 11)*

A few comments are in order. Lysandrou does not suggest that money is a commodity with intrinsic value, simply that the basic unit of "money" is something that can be traded impersonally. But money is traded impersonally. We hope that we have made clear throughout our discussion that the nature of money should not be confused with its generation and its allocation. Money as bank credit is in itself traded at par. If an entity receives a payment, it does not know if the money comes from a loan or from the central bank's QE. On the other hand, as stated in Mehrling (2015), money is intrinsically hierarchical. Bonds need base or credit money. A bond is by nature a debt while money can effectively be thought of as purchasing power.

This being said, the aggregation of principal players in firms, households, banks, and financial markets is misleading. The principal players are ultimately households and capitalists. The other entities are human artefacts. Capitalists are households too but they own firms, either directly or indirectly through professional asset managers. Of course in the real world things are not so clear-cut: many households own stocks. In fact, in the U.S. an estimated 90 million persons[5] out of a population of 325.7 million own stocks.

Box 8.2 The Evolution of the Money Circuit

To gain a better understanding of the evolution of the money circuit let's perform a didactical thought experiment. Assume that in an idealized closed economy the circuit of money is exactly as described by Circuitism: households receive wages but do not borrow, firms borrow to pay wages (initial finance) and collect money when they sell their products to households so that they can repay their loans (final finance). We also observed that capitalists can borrow money and buy the extra production not bought by workers. These purchases create profits for the firms which can then distribute these profits to the owners of the firms (i.e. capitalists) as dividends; the capitalists can then repay their bank debt.

Now suppose that at a given moment households borrow money to pay for new homes and new cars. As originally argued by Wicksell, this borrowed money adds to the purchasing power of households that can thus buy the homes and cars. In aggregate the money used to buy homes and cars becomes profits for the firms that provide these items. This profit will be retained by firms or distributed to capitalists as dividends. Capitalists can use the money obtained through dividends to buy the extra products and services that workers cannot afford. Firms can use retained profits to finance production, thus reducing their borrowing.

Of course households must repay debt with their wages. Suppose that at every period households take out new loans so that the total outstanding debt remains constant. In this situation, the borrowing of households has created a stock of money that circulates between firms and capitalists. This money in circulation allows firms to make monetary profits and pay dividends. It does not really matter if money is kept as retained profits or immediately paid out as dividends. In fact, if money is used as retained profits, firms will use it to reduce their borrowing in order to make a profit. This new circuit of money is a bit paradoxical: households borrow money – at a cost of course – and capitalists reduce their borrowing and make a profit.

To continue with our thought experiment, suppose that households increase their borrowing. More money circulates and firms can make larger monetary profits and distribute bigger dividends. Progressively, as households increase their borrowing firms will reduce theirs and finally all borrowing will be made by households. The upper limits of household borrowing are not immediately obvious as loans can be granted on the basis of collateral, that is, items already

acquitted such as their homes whose value keeps on going up due to the flow of money.

To continue our thought experiment we can now introduce financial assets. The most obvious assets are corporate stocks. Capitalists can use their money to trade stocks. The trading of stocks has two main effects: it shifts ownership among capitalists and eventually changes the stock market capitalization as prices depend on trading. In addition to stocks, modern financial engineering has created a vast array of securities that are essentially packages of debt. Capitalists can use their money to trade financial assets based on debt. The trading of debt-based assets creates a complicated web of debt/credit relationships but does not create money. This is because packages of debts cannot be used as means of payments. Financial assets, even very liquid ones, such as money market funds, are used as a storage of value.

We can see from our thought experiment that the circuit of money has changed from an idealized circuit centered on firms, where only firms borrow money, to a realistic circuit where households create a stock of money used by firms and capitalists. The sustainability of this process is clearly questionable as it is ultimately based on the ability of households to repay their load of debts. However, the trading of financial markets has created additional loops where the growing value of financial assets allows firms and capitalists to borrow money for speculative reasons in many ways other than taking out loans. In addition, the Great Recession following the 2007–2008 market crash pushed central banks to step in with quantitative easing, buying assets and thereby creating even more money.

The present situation is that of a large pool of money created by households and used by firms and capitalists, a large pool of money created by central banks with quantitative easing and used primarily for trading assets, and another pool of money, difficult to quantify, used for speculative purposes.

8.3 Puzzles: the declining velocity of money and missing inflation

The dynamic theory of the circuit of money discussed earlier provides a natural theoretical underpinning for studying the relationship between money and the real economy. But before proposing a modelling macroeconomic framework, let's discuss two recent puzzles in monetary finance: the puzzle of the shrinking velocity of the circulation of money and the puzzle of missing inflation.

The velocity of circulation of money is a measure of how fast money circulates in the economy. It is measured by the ratio of the nominal GDP to the money stock. There are different velocities for different money stocks. Exhibit 8.6 presents data for the U.S. economy.

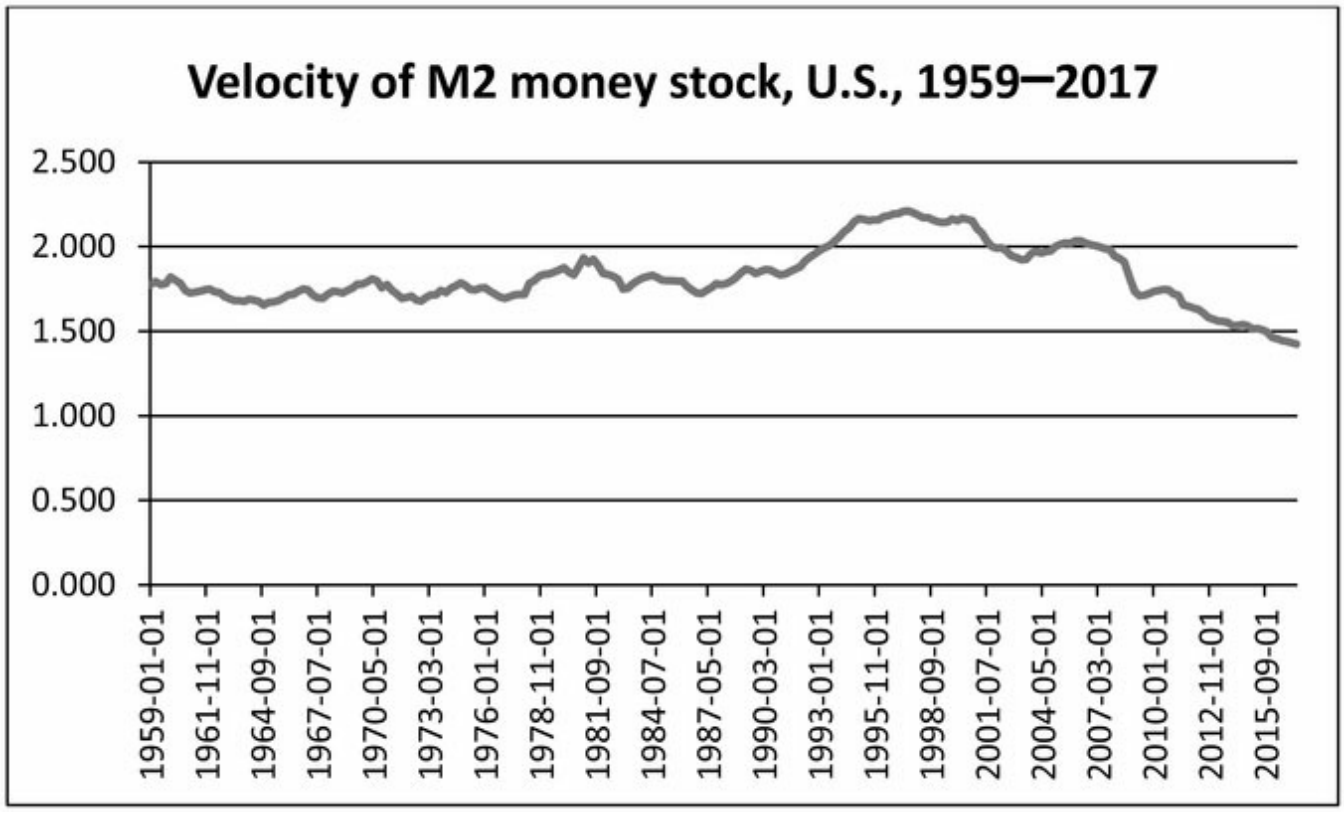

EXHIBIT 8.6 The velocity of M2 money stock in the U.S. for the 58-year period from 1959 to 2017. Constructed by the author based on data from FRED (Federal Reserve Economic Data) from the Federal Reserve Bank of Saint Louis.

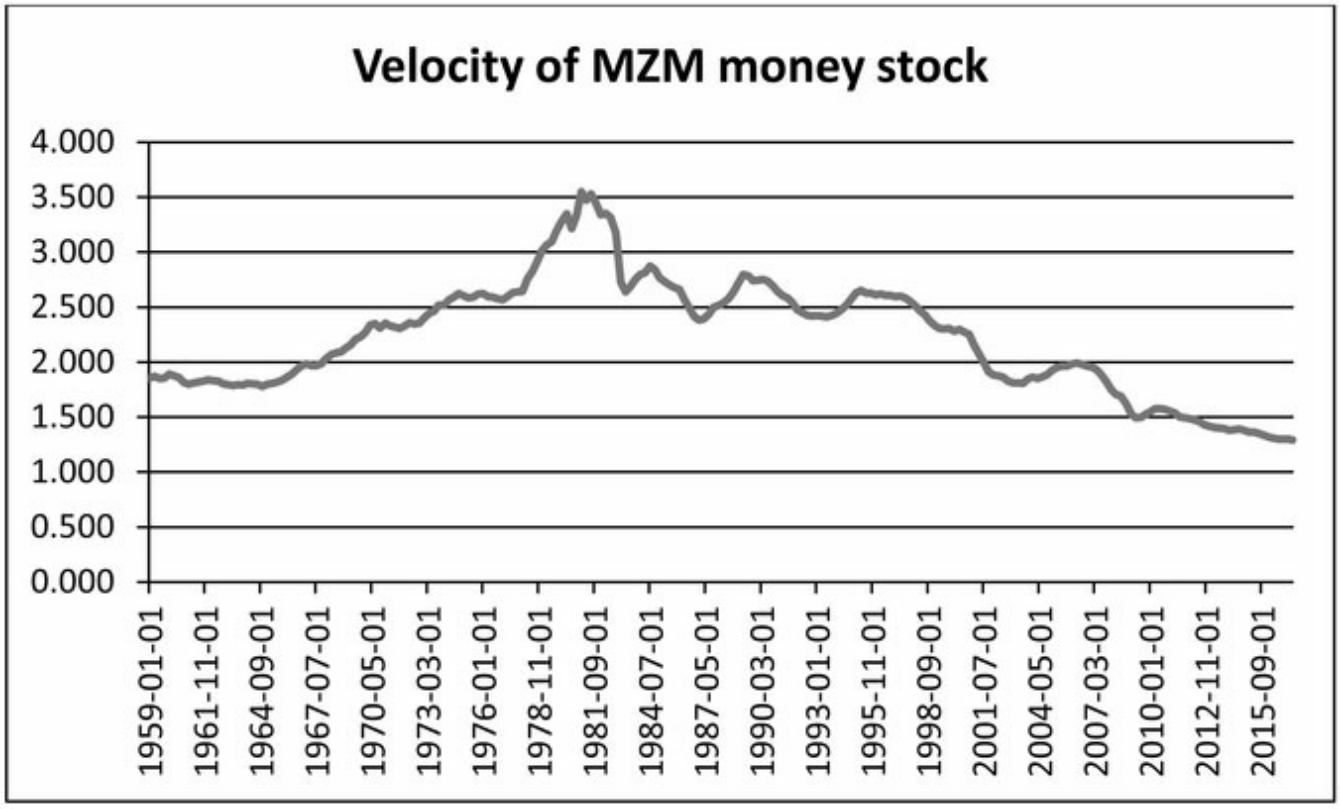

EXHIBIT 8.7 Velocity of MZM (money with zero maturity) money stock for the U.S. in the 58-year period 1959–2017. Constructed by the author based on data from FRED (Federal Reserve Economic Data) from the Federal Reserve Bank of Saint Louis.

As can be seen from Exhibit 8.6, over the period 1959 to 1990, the GDP/M2 ratio was stable, oscillating around a value of 1.7. The ratio reached roughly 2.2 for a short period but then started going down and is now around 1.4. Exhibit 8.7 shows the velocity of U.S. MZM (money with zero maturity) money stock in the period 1959–2017.

The behavior of the ratio GDP/MZM is slightly different. The ratio grew almost steadily from 1959 to reach a peak of 3.5 by 1980, and then steadily declined to the current value around 1.4.

Recall that the Federal Reserve Bank of Saint Louis defines the money stocks as follows:

> M1 is the money supply of currency in circulation (notes and coins, traveler's checks [non-bank issuers], demand deposits, and checkable deposits). The broader M2 component includes M1 plus saving deposits certificates of deposit (less than $100,000), and money market deposits for individuals. MZM (money with zero maturity) is the broadest component and consists of the supply of financial assets redeemable at par on demand: notes and coins in circulation, traveler's checks (non-bank issuers), demand deposits, other checkable deposits, savings deposits, and all money market funds.

The decline of the velocity of money represents a theoretical puzzle when considered together with the evolution of the inflation rate (as currently measured

through the Laspeyres or Paasche index on a basket of goods and services). In fact, observe the inflation data for the U.S. presented in Exhibit 8.8.

EXHIBIT 8.8 U.S. inflation rate for the 56-year period from 1960 to 2016. Constructed by the author based on data from FRED (Federal Reserve Economic Data) from the Federal Reserve Bank of Saint Louis.

From Exhibit 8.8 we can see that inflation grew rapidly during the period 1960–1980 from less than 2% to a peak of almost 14% before declining abruptly, to fluctuate thereafter between 4% and 0%, reaching zero or negative values twice. The decrease in the velocity of money signifies that the money stock expands more rapidly than the growth of GDP. But according to mainstream economics, an expanding monetary mass should produce inflation. So why hasn't inflation grown? Up to 1980, the velocity of M2 remained approximately constant with only a temporary increase around the last decade of the 20th century while inflation increased rapidly. In the last two decades, inflation has remained low and velocity declined.

Similar behavior has been observed in most developed economies and in particular in Japan. To explain the "puzzle" of the declining velocity of the circulation of money, Friedrich Werner (1992, 1997) proposed an explanation: the Quantity Theory of Credit (QTC). The QTC divides the total credit available, Q, and therefore the credit money available, into two categories: Q(gdp), money which is used for commercial and industrial transactions; and Q(assets), which is money used for purchasing assets. Only Q(gdp) contributes to actual productive and commercial transactions and therefore only Q(gdp) should be used to measure the velocity of money. Using empirical data from Japan, Werner showed that this decomposition explains the puzzle of velocity.

The same problem can be looked at from a slightly different angle. Daniel Gros (2017), director of the Brussels-based Center for European Policy Studies, writes:

> Central banks have a problem: growth in much of the world is accelerating, but inflation has failed to take off. Of course, for most people, growth without inflation is the ideal combination. But central banks have set the goal of achieving an inflation rate of "below, but close to 2%," as the European Central Bank puts it.

Gros observes that though trillions of dollars and euros have been injected into the world economy through QE and though interest rates are near zero, inflation has not picked up and wages are not growing, even in countries like the U.S., Germany, and Japan where measured unemployment is low. The reasons for the missing inflation are not generally recognized. One could expand the reasoning of Werner's QTC and observe that inflation is perhaps not picking up because the money created through QE has not reached the real economy but has remained in financial markets, producing asset inflation but no inflation in goods and services.

This explanation is only partially true. Ultimately not much money has reached the real economy. But it seems clear that something is missing in the explanation: economies in the U.S. and in Europe are nominally growing at modest but still positive rates. We suggest that puzzles such as the declining velocity of money and missing inflation coupled with a rapid expansion of the money stock cannot be explained without considering that modern economies are complex systems. This is the subject of the next section.

8.4 Money and complexity

Economies are systems formed by many interacting agents and are therefore a natural subject of investigation for the theory of complex systems. A key question is whether the macroscopic behavior of economies can be explained in terms of the behavior of their constituents. This question was implicitly present in Adam Smith's notion of the "invisible hand" that coordinates market activities. Following Lucas, neoclassical economics again attempted to base macroeconomics on agents' behavior (i.e. microeconomics) but avoided the issue of complexity by postulating that agents have no mutual interaction and behave as mutually independent utility optimizers.

The study of economics as a topic in the theory of complex systems is associated with The Santa Fe Institute, the first research institute devoted to the study of complexity. *The Economy as an Evolving Complex System I* edited by Philip Anderson (1988), recipient of the 1977 Nobel Prize in Physics, and *The Economy as an Evolving Complex System II* edited by Bryan Arthur (1999), are collections of seminal papers on economics and complexity. Beinhocker (2007), mentioned earlier, offers a non-technical introduction to economics and complexity. The development of multiagent systems[6] and languages to describe multiagent systems has given additional impulse to the study of economic complexity.

The study of economics and complexity has focused on a number of key topics, among them the study of the aggregation and clustering of economic agents, networks, and self-organizing behavior. This line of investigation has allowed researchers to explain mathematically the emergence of the distribution of clusters and aggregates which, in turn, explains why many economic variables – from financial returns to the household income – have inverse power law distributions.[7] In fact, the theory of random interacting structures such as random graphs and random networks shows that the distribution of the size of clusters follows an inverse power law. The theory of networks has made important contributions to the study of the instability of financial systems. Self-organization, that is, the ability of complex systems to self-organize in aggregates characterized by meaningful behavior, is also a key to the study of economic complexity. The objective is to explain how complex social structures such as ancient empires emerged autonomously from sparse bands of hunters.

Self-organization implies an increase in information and a reduction of disorder. Hidalgo (2015) offers a non-technical introduction to the scientific issues of the growth of information, explaining how order can increase locally using exogenous sources of energy. On the practical side, it has been argued that economic prosperity is intrinsically related to complexity. At MIT Hausmann et al. (2014) created

measures of economic complexity, in terms of both products and economic relationships. Ranking the nations in terms of complexity, they found that those nations that rank highest in complexity are, perhaps not surprisingly, those ranked highest in terms of the rate of economic growth.

A different perspective on the theory of complexity – one which we will explore – might investigate how complexity affects the interaction of money with the economy. In a nutshell, in modern, highly complex, rapidly evolving economies, concepts such as real GDP and real growth are ill-defined. GDP and growth are increasingly determined by the money supply and by financial structures – without the possibility of discriminating real from nominal quantities. In a highly complex society the global level of economic activity as measured by GDP is somewhat conventional.

8.4.1 Economic growth in advanced societies

Economic growth is a topic of great importance in modern economics; it impacts the quality of life of a nation's citizens. Lucas is credited with having said that after having thought about growth it was difficult to think about anything else. The ability to produce economic growth has separated modern developed countries from countries that have not been able to develop at the same pace, though clearly in some instances, the gap is closing. Angus Deaton (2013), recipient of the 2016 Nobel Memorial Prize in Economic Sciences, analyzed the characteristics that mark the path of growth of modern economies. His list includes the ability to grow crops and produce food, breed livestock, build homes and secure clean water, sanitization, transportation, and medical treatment. Growth then has a concrete meaning that can be measured by the aforementioned parameters: escape from poverty is based on satisfying basic needs.

In modern complex economies basic needs such as clean water, food, and warm shelter are generally satisfied though there are still genuine improvements to be made, for example, to reduce poverty, malnutrition, poor housing, and a degraded natural and social environment. In many cases the entire economy would benefit from better infrastructure. From the point of view of the offering of products and services, there are areas where improvements are possible and needed. The development of modern medicine is one area where genuine improvements are still being made. However, there are fewer and fewer areas where technology can satisfy essential needs that would be shared by the entire population.

8.4.2 How do we measure economic output?

Standard macroeconomic textbooks such as Mankiw (2012) posit that an economy produces a composite good. However, no physical measure of the composite good is possible: we simply cannot aggregate automobiles, laptops, cruises, and all other products. Aggregation is made through economic value, represented by price. Aggregation by price produces the nominal output, that is the output aggregated using current prices. In order to compare output in different moments, and therefore compute growth, mainstream economics posits that the nominal output must be deflated with a price index to yield the real output. Real growth is the change in real quantities.

This procedure might give an approximate account of the evolution of an economy if products are stable over most of product space and if growth is essentially quantitative. For example, after World War II, in most European countries, growth consisted in rebuilding homes and infrastructure, manufacturing automobiles and home appliances, and providing services such as medical services and leisure. In this case, the dichotomy of real vs. nominal quantities has a reasonably clear meaning and economic growth is rooted in the numerical growth of physical parameters such as the number of cars manufactured.

Observe that measuring real GDP and real growth is not an approximation of a true GDP and a true growth rate because there is no such thing as a true GDP and a true growth rate. We can say that operationally we define a number of variables, such as nominal GDP, observed inflation rates, and other macro variables, and try to find relationships with predictive character. This is the only test of the validity of our macroeconomics: macro variables are meaningful only if they (1) are operationally defined and (2) can be placed in a set of relationships with a predictive character, that is, forecast the behavior of the object of our study. We find once again an instance of the methodological principle that science is not the refinement of intuition. Any household has a simple intuition of growth: one moves to a bigger house, drives a more powerful car, wears designer clothes . . . However, these changes in themselves lose value if friends and neighbors experience the same change. From the point of view of a household, growth is a mix of genuine physical improvements and improvements relative to others.

The above framework of nominal and real quantities does not reflect the current economic reality of advanced economies. The output of modern advanced economies is formed by a staggering number of products, estimated by Beinhocker (2007) to be in the range of hundreds of millions. And a good fraction of the estimated hundreds of millions of products change continuously.

Growth today is primarily qualitative, an increase in complexity. For example, in most advanced countries the number of automobiles per capita is not growing. Exhibit 8.9 illustrates the evolution of the number of automobiles per thousand inhabitants in the U.S. in the 68-year period from 1946 to 2014. While the number of cars grew rapidly and almost linearly from 1946 to 1990, growth then flattened and is now oscillating around 800 cars per 1,000 inhabitants. While the number of cars manufactured has virtually stabilized, new features are continuously being added, such as the automatic control of driving parameters or automatic braking in the case of an emergency. Performance in terms of horsepower has been pushed beyond the ability to actually use it.

The relationship between the price and features of products is blurred. Prices might change significantly due to factors other than product features. This is obvious in sectors such as fashion, where the market value of products depends on what celebrity is advertised using them and might change in function of a personal event of the same celebrity, such as marital infidelity. The market value of products depends not only on the product itself and its image but also on a volatile web of relationships with other products. This is clear in electronics where the value of a software application depends on its ability to interact with many other (popular) applications.

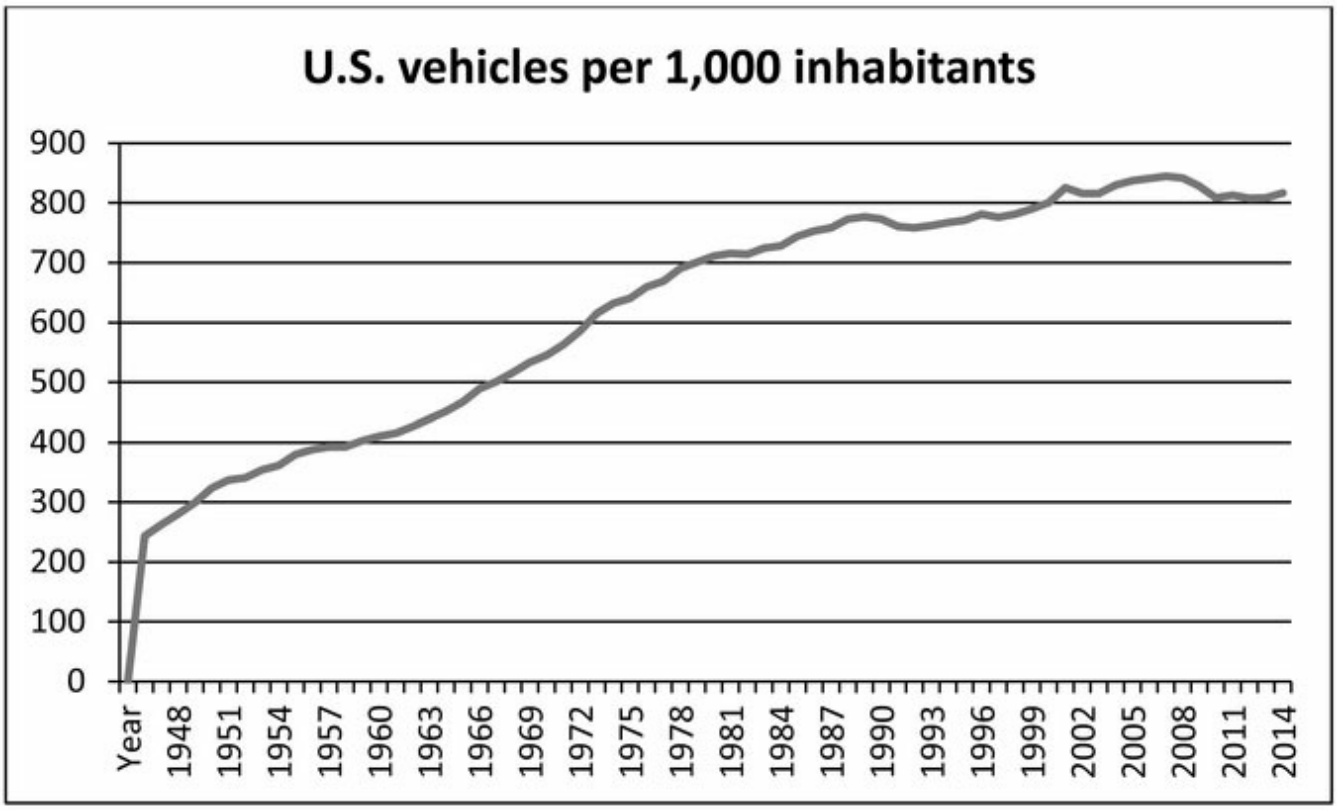

EXHIBIT 8.9 The evolution of the number of automobiles per thousand inhabitants in the U.S. in the 68-year period 1946–2014. Constructed by the author using data from table 3.6 of the Transportation Energy Data Book of Oak Ridge National Laboratory.

The classical framework of computing a nominal GDP and then deflating it to obtain the real GDP is not useful when applied to today's complex economies. As discussed in Chapter 6, price inflation is not a well-defined concept in today's complex, rapidly evolving economies. What indices such as the Laspeyres or Paasche indices compute is the change of value of a panel of products consumed by low- to middle-income households. In other words, widely used inflation indices are applicable only to the most static part of the economy.

A different framework of analysis is required for complex, evolving economies. The first step consists in understanding the role of money in forming the GDP. We discussed earlier the circuit of money from initial to final finance. In order to purchase products agents need money. In the classical view, if too much money is available, prices increase and inflation appears. Therefore, as soon as inflation above the target rate is detected, central banks reduce the money supply.

But this is not what happens in modern economies. First, following Werner's QTC, the total amount of money available to create purchasing power must be divided into two parts. The first, Q(gdp) part contributes to GDP-related transactions while the other part Q(assets) is only used to trade assets. This is what happened with QE that injected a large amount of money, above $4 trillion in the U.S. alone, that remained confined to financial markets thereby creating asset inflation. The split of money creation into two parts, one of which stays in financial markets and the other of which reaches the real economy, explains the decline in the velocity of money as well as why financial markets can grow at a much higher speed than the real economy.

What about the real economy? The Q(gdp) is not distributed uniformly to all economic agents. Actually it is distributed very unequally. The largest percentage of the population lives on wages. We discussed earlier how the balance of power has shifted to firms and capitalists, and profits are now responsible for a much larger share of consumption.

While the portion of the population that earns wages has seen their income stagnate or decrease, the portion of the population that earns profits has seen their income increase significantly, thanks to capital gains and the payments of dividends. The effects are twofold. First there are sectors of the population whose income grows much more rapidly than the rest of the economy. This is demonstrated by the increase in income inequality in many Western economies. Second, this income is typically spent on products and services characterized by a high level of complexity and innovation with a snob element. It is basically impossible to measure inflation in these product segments. The end result is that the bulk of the economy sees no increase in its purchasing power, does not grow, and does not exhibit inflation as

currently measured. Nominal growth is concentrated in small subsectors of the economy.

8.4.3 Macroeconomics in the age of complexity

The above considerations explain the current macroeconomic puzzles which we can restate as follows: how is it that the monetary mass has grown significantly, the velocity of circulation has declined, market capitalization is way up, and modest economic growth is coupled with virtually no inflation?

We suggest that the explanation can be found in the complex nature of modern economies. Nineteenth-century economies were limited by production capabilities. The affluent portion of the population appropriated a large fraction of the scarce output available. Industrial automation progressively made available more physical output to larger fractions of the population. For example, as seen in Exhibit 8.9 the number of automobiles per thousand U.S. inhabitants went from 250 in 1960 to 700 in 1980. Progressively products such as automobiles and services such as telecommunications have become more complex, more sophisticated.

This process has accelerated rapidly with the digital revolution and other technological advances. Today economies are less limited by production constraints as both demand and supply have shifted towards complexity. Economic value is often created artificially by social constraints. The key constraint in modern economies is profit. Wages are constrained by firms' objectives to increase profit. The bulk of the economy finds limits to growth as wages are stagnant and private debts cannot grow indefinitely. Growth is concentrated in small subsectors of the economy.

Complexity and the rate of innovation render impossible the use of inflation measurement to measure growth. Growth is determined to a large extent by purely financial considerations. Growth theory needs to be revised. Economies are becoming increasingly virtual in the sense that the commercial value of products is delinked from their characteristics and increasingly linked to a web of social considerations including one's purchasing power.

8.5 Economic and financial instabilities

Hyman Minsky was perhaps the first to realize that a capitalistic economy with a monetary system is subject to instabilities. Mainstream economics claims that, if market processes are left undisturbed by the government, capitalistic economies are self-stabilizing systems that keep economies in a situation of equilibrium. As observed earlier in this chapter, mainstream macroeconomic models do not include financial systems because financial systems might create situations of disequilibrium.

As Minsky argued that capitalistic economies are not self-stabilizing systems but are prone to crises, he was considered outside of the mainstream and his work was largely ignored. It was only in the aftermath of the 2007–2008 financial crisis that his work was rediscovered and received attention.

Minsky's reasoning is simple. Consider an economy which has experienced a period of prosperity. Progressively businesses become more confident and are willing to undertake projects that were previously considered too risky. The climate of confidence extends to bankers who begin to accept proposals for loans that were previously considered too risky. The economy expands, the amount of debt grows, and stock market capitalization increases. The process continues until businesses begin to experience cash-flow problems. At this point, there is a rapid fall in confidence and in the level of business activity; a serious financial crisis might follow.

The 2007 U.S. subprime mortgage crisis provides an illustration of Minsky's theory. Subprime mortgages were granted on the basis that the value of property would increase (of course there were other considerations such as the fact that banks extended credit to persons who could not be considered creditworthy). Mortgages were repackaged in asset-backed securities and traded. When mortgages could not be repaid (perhaps unsurprisingly), the crisis started and quickly spread throughout an overleveraged economy and the world as investors worldwide had bought the securities.

Do similar considerations apply to our framework? We have seen how an excess of money creation is shared between small sectors of the real economy and financial markets. This flow of money produces fast nominal growth of financial markets and fast nominal growth of a subset of the economy. Both grow at rates much higher than the overall economy. Is this situation sustainable?

The answer depends on a number of factors – one reason why it is difficult to predict crises. On one side it is clear that growing money flows imply a growing amount of debt; in this sense, the present situation is vulnerable to Minsky's analysis.

However, we also see growing profits in some sectors of the economy. Some of these profits come from subsets of the economy that output complex products and services, where detecting inflation is difficult. Stock-flow consistent modelling of economies as multiagent systems offers one way to model this situation; we will shortly discuss stock-flow consistent modelling but let's first look at Modern Money Theory.

8.6 Modern Money Theory

Modern Money Theory (MMT) is a macroeconomic theory that builds on the work o many economic thinkers: Knapp and the Chartalists, the Circuitists, Keynes Schumpeter, Abba Lerner's functional finance, Wynne Godley's notion of stock-flow consistency, and Minsky's analysis of economic instability.

Early statements of MMT are due to Warren Mosler, an American entrepreneur and economist, who first presented the ideas of MMT in two short books: *The Seven Deadly Innocent Frauds of Economic Policy*(2010) and *Soft Currency Economics II* (2013).[8] Bill Mitchell, professor of economics and director of the Centre of Full Employment and Equity (CofFEE) at the University of Newcastle, New South Wales is also an early representative of MMT.[9] Other prominent exponents of MMT include Steve Keen, an Australian-born economist, Pavlina Tcherneva at Bard College (New York) and the Levy Economics Institute, and Randall Wray at the University of Missouri–Kansas City, research director of the Center for Full Employment and Price Stability and scholar at the Levy Economics Institute of Bard College.

MMT integrates both Knapp's verticalism and Moore's horizontalism (discussed in Chapter 5). MMT claims that money is immaterial fiat money, created by the authority of the state and endogenously expanded by commercial banks. MMT's original contributions are primarily in the area of fiscal policy: MMT accepts Lerner's ideas of functional finance, that is, MMT accepts the principle that economic policies of sovereign states should be dictated by economic, functional reality and not constrained by monetary considerations.

MMT proponents maintain that sovereign states can generate all the money they need to implement their policies. They also maintain that taxation and the issuing of state debt are not necessary for financing the state but are suggested by considerations of price stability. In the writings of MMT this point is often expressed in ways that might easily lead to a rejection. Let's now discuss the reasoning of MMT.

MMT suggests considering a single entity that includes both the Treasury and the central bank. It does not propose the actual fusion of the two but only to think of them as a single entity. This combined entity, a sort of conceptual black box, orders goods and services for the state, pays the salaries of state employees, and receives taxes and the proceeds of the sale of debt if any debt is issued. That is, MMT proposes looking at the functioning of the Treasury and the central bank globally without looking at their internal relationships.

The first consideration made by MMT is that people and firms must first receive money and then pay taxes. Therefore the Treasury/central bank must create and

advance the money needed for the functioning of the state. In this sense, the state is like any firm which first must finance production and then can sell products. Citizens can only pay back to the government what they have received. Note that this is contrary to prevailing practice in most modern economies, including the U.S., the U.K., and the European Union, where central banks cannot finance the state directly.

The main tenet of MMT – that the government can create and advance any amount of money – implies that money can be created by the central bank in the same way that commercial banks create money: the central bank simply credits the deposit of the government. The fact that the central bank also registers a debt from the government is, for all practical effects, irrelevant. This money then circulates in the economy. As most citizens and firms pay taxes, the money advanced by the government reaches successively the entire economy; were this not the case, some people would run a permanent deficit with the state.

MMT's next point is that the government imposes taxes not to finance its operations but to ensure price stability. This point can easily become controversial if only due to its interpretation. The idea is that the government could in principle continue to order goods and services by issuing new money. MMT does not propose to finance the state by issuing new money but holds that there are no intrinsic limits to the process of the self-financing of the government. The government taxes not because it needs the money but to withdraw money from the public as a continuous accumulation of money would lead to price instability. Much of the controversy around this proposal stems from the unconventional way of stating it. MMT maintains that in principle the government can function issuing new money.

Assume that the government taxes citizens and firms. MMT sees a problem here: How can people save? Endogenous money generation by commercial banks could satisfy the need of a production economy as was explained by Graziani. Money generated by commercial banks can be thought of as needed for industrial or commercial activities. This money has a time-limited function, that is, it exists until products or services are sold and money can thus be destroyed. This money cannot really serve the need of savers as the savings of some people or firms imply that other people or firms pay interest on their loans simply to create money that disappears in savings.

The government, therefore, must choose a taxation level that does not withdraw all the money that has been advanced but only a fraction of it. The remaining fraction is what the public can save. In other words, the government must run a deficit in order to allow people to save. This is a tenet of MMT which, understandably, has stirred controversy.

A budget deficit could be covered by simply issuing new money. The money

issued by the government to cover deficit would end up as the savings of people and eventually firms. Savings would become bank deposits. But in this situation deposits would not earn any interest. In fact, what entity would be willing and able to pay interest on savings?

MMT suggests that the government issue debts to give savers a risk-free source of interest. Interest can be financed by issuing new debt but clearly in the long run the cost of paying interest must be borne by all citizens.

Note that the decision to save is a decision made by individuals, while the decision to run a deficit is a government decision. If people do not save, they have no money set aside for future consumption but spend all their money on current consumption. Deciding to save, individuals have several options. They can open bank deposits, earning virtually no interest; they can buy riskless government securities, an action that returns the money to the government; or they can buy securities issued by private entities. In the latter case money keeps circulating in the economy. Of course citizens can also use their savings to buy permanent goods such as houses.

Clearly at this point political as well as economic considerations play a role. MMT rejects the notion that there are sound financial justifications for austerity measures. Following Abba Lerner, MMT claims that budgets should be defined in real terms, considering the real needs of the economy and, of course, the real resources available. Lerner considers financing needs to be a technical question.

8.7 Stock-Flow Consistent models

Stock-flow consistent (SFC) models of production economies with a financial system are non-mainstream models able to capture non-equilibrium features of modern economies. Two types of models have been proposed: analytical stock-flow consistent models and agent-based stock-flow consistent models. There is ample literature on both. Implicit in Copeland's flow of funds (1949, 1952), the SFC approach is rigorously constructed and described in Godley and Lavoie (2007). Caverzasi and Godin (2015) review the literature on SFC models and Caiani et al (2016) review the literature on agent-based SFC and propose a benchmark agent-based SFC model. Tymoigne (2006) discusses in detail the Minskyan framework and provides a SFC Minskyan model. Current SFC models are typically written in some system dynamic language that automatically generates equations from a high-level description of the system.

A SFC model has three main components: the accounting matrix, the balance sheet, and the dynamic equations. The accounting matrix determines the constraints that represent stock-flow consistency. The dynamic equations determine the dynamic evolution of the system. A realistic SFC model can be quite complex. Here we can only give an idea of the logic of a SFC model.

To illustrate how SFC models are constructed, let's describe the simplest model in Godley and Lavoie (2007), the model SIM. Model SIM describes a closed economy composed of three sectors: households, firms, and a government. There are N households in the system. Money is created by the government as fiat money:

- Households receive wages W in exchange for labor, pay taxes T, and consume C out of their disposable income YD;
- Firms produce an output Y, which is sold to households and the government, and pay wages in exchange for labor;
- Government buys output G from the firms and receives taxes from the household sector.

There is only one asset, the money stock H. Therefore, income that is not consumed is saved as cash. Of course, if households have positive savings, then the government runs a deficit. The model is described by the following equations:

$YD = W * N - T$: the disposable income of households is equal to total wages minus taxes.

$T = \theta * W * N$: taxes are proportional to total wages.

$C = \alpha\ YD + \beta H_{h,-1}$: consumption is the sum of a fraction of disposable income plus interest.

$Y = C + G$: output is equal to consumption plus government spending.
$N = Y/W$: the number of households is equal to output divided by individual wage.
$\Delta H_s = G - T$: changes in the government deficit are equal to expenditure minus taxes.
$\Delta H_h = YD - C$: changes in household savings are equal to disposable income minus consumption.
$\Delta H_s = \Delta H_n$: changes in government deficits are equal to changes in household savings.

The transaction flow matrix is shown in Table 8.1 and the balance sheet in Table 8.2.

Though clearly unrealistic, this model gives an idea of the nature of SFC models A realistic model is described in Tymoigne (2006) where a banking system is added to the economy.

Agent-based SFC models (ABSFC models) are by nature more complex as th behavior of agents and the markets in which they operate needs to be described. Caiani et al. (2016) provide a general description of an ABSFC model. Their intent

TABLE 8.1 Transaction flow matrix

	Households	*Production*	*Government*	*Sum*
Consumption	−C	+C		0
Government expenditures		+G	−G	0
Wages	+W*N	−W*N		0
Taxes	−T		+T	0
Change in money stock	−ΔH		+ΔH	0
Sum	0	0	0	0

TABLE 8.2 Balance sheet

	Households	*Production*	*Government*	*Sum*
Money stock	−ΔH	+C	+ΔH	0

is to provide a benchmark model to satisfy the research needs of scholars. Thus far the only large-scale implementation of ABSFC models is the Eurace project of the European Union.[10]

Let's now briefly describe how an ABSFC model can include the features discussed previously. A model should have three features that are absent in models developed thus far:

1. A segmentation of the economy should be introduced so that money can flow

unequally in different segments of the economy;

2. Two types of production should be introduced: the production of goods and the production of complexity. This is an idealization of the fact that product innovation brings more complex products. The production functions of goods and of complexity are different: the production of complexity has a higher level of elasticity than the production of goods;
3. A measurement of inflation should be explicitly introduced in the model with appropriate feedback. The objective here is to simulate how different growth paths inside the same economy – including fast growth of asset values – might coexist with very low inflation as usually measured, and with a large increase in the money stock.

Notes

1 The Institute for New Economic Thinking (INET) is a not-for-profit think tank whose purpose is to suppo academic research and teaching in economics "outside the dominant paradigms of efficient markets and rational expectations." Founded in 2009 with the financial support of George Soros, INET is a response t the global financial crisis that started in 2007. For more information, see http://ineteconomics.org/

2 Recall from Chapter 1 the study by Fichtner, Heemskert, and Garcia-Bernardo (2017) that mapped th ownership of the "Big Three" passive index funds in the U.S. – BlackRock, State Street, and Vanguard · and found that they constitute the largest shareholder in 40% of all listed U.S. corporations and 88% of the firms in the S&P 500 index.

3 System Dynamics is an approach to studying the non-linear behavior of systems made up of multiple interacting parts. Developed by Jay Forrester at MIT, system dynamics has originated a number o languages that describe complex dynamic systems using simple building blocks and automatically generate equations.

4 A strict application of Circuitism was not possible even in the 1950s and 1960s as households borrowed money to finance the purchase of their homes.

5 The figure comes from Investment Company Institute (2016). A 2017 Gallup poll found that for the perio 2009–2017, 54% of Americans said they owned stocks, down from 62% for the period 2001–2008 (see Jones 2017).

6 Multiagent systems in economics and finance are software programs made up of a collection of sub-programs that simulate the behavior of real agents and interact in some artificial market.

7 An inverse power law is a probability distribution that decays as a power law with a negative exponent.

8 Mosler maintains a website (www.moslereconomics.com) where one can find many resources and opinions on MMT.

9 Mitchell keeps a blog: http://bilbo.economicoutlook.net/blog/

10 The Eurace model is computationally intensive and requires large-scale parallel computers to run properly. The Eurace project is described in the official site www.eurace.org and in a number of papers. See for example Deissenberg et al. (2008) for a description of the project and Raberto et al. (2012) for a discussion of some computational experiments related to building an ABSFC model.

References

Anderson, Philip W., Kenneth Arrow, and David Pines, editors. 1988. *The Economy as an Evolving Complex System I.* Santa Fe Institute Series. Boulder, CO: Westview Press.

Arthur, W. Brian, Steven N. Durlauf, and Davi Lane, editors. 1999. *The Economy as an Evolving Complex System II.* Santa Fe Institute Series. Reading, MA: Addison-Wesley.

Barro, Robert. 2009. "Lucas Roundtable: Don't Fault the Economists." *The Economist*. (August 8).

Beinhocker, Eric D. 2007. *The Origin of Wealth; The Radical Remaking of Economics and What It Means for Business and Society*. Boston, MA: Harvard Business Review Press.

Bruun, Charlotte and Carsten Heyn-Johnsen. 2009. "The Paradox of Monetary Profits: An Obstacle to Understanding Financial and Economic Crisis?" Discussion Paper No. 2009–52. (December 3): www.economics-ejournal.org/economics/discussionpapers/2009–52.

Caiani, Alessandro, Antoine Godin, Eugenio Caverzasi, Mauro Gallegati, Stephen Kinsella, and Joseph E. Stiglitz. 2016. "Agent Based-Stock Flow Consistent Macroeconomics: Towards a Benchmark Model." *Journal of Economic Dynamics and Control*, vol. 69:375–408.

Cauwels, Peter and Didier Sornette. 2012. "The Illusion of the Perpetual Money Machine." The Swiss Financial Institute Research Paper, No. 12–40.

Caverzasi, Eugenio and Antoine Godin. 2015. "Post-Keynesian Stock-flow-consistent Modelling: A Survey." *Cambridge Journal of Economics*, vol. 39, no. 1 (January):157–187.

Copeland, Morris. 1949. "Social Accounting for Moneyflows." *The Accounting Review*, vol. 24, no. 3:254–264.

Copeland, Morris. 1952. *A Study in Moneyflows in the United States*. Cambridge, MA: National Bureau of Economic Research/ Arno Press.

Deaton, Angus. 2013. *The Great Escape: Health, Wealth, and the Origins of Inequality*. Princeton, NJ: Princeton University Press.

Deissenberg, Christophe, Sander van der Hoog, and Herbert Dawid. 2008. "EURACE: A Massively Parallel Agent-based Model of the European Economy." Applied Mathematics and Computation, vol. 204, no. 2 (October 15):541–552.

Fichtner, Jan, Eelke M. Heemskert, and Javier Garcia-Bernardo. 2017. "Hidden

Power of the Big Three? Passive Index Funds, Re-Concentration of Corporate Ownership, and New Financial Risk." *Business and Politics*, vol. 19, no. 2 (June):298–326.

Godley, Winne and Marc Lavoie. 2007. *Monetary Economics: An Integrated Approach to Credit, Money, Income, Production and Wealth.* New York: Palgrave Macmillan.

Graziani, Augusto. 1990. "The Theory of the Monetary Circuit." *Economies et Societies*, vol. 24, no. 6:7–36.

Gros, Daniel. 2017. "The Inflation Target Trap." *Project Syndicate.* (July 6): www.project-syndicate.org/commentary/inflation-targeting-risks-by-daniel-gros-2017–07.

Hausmann, Ricardo, César A. Hidalgo, Sebastián Bustos, Michele Coscia, Sarah Chung, Juan Jimenez, Alexander Simoes, and Muhammed A. Yıldırım. 2014. *The Atlas of Economic Complexity: Mapping Paths to Prosperity*. Cambridge, MA: MIT Press.

Hidalgo, César. 2015. *Why Information Grows: The Evolution of Order, from Atoms to Economies*. New York: Basic Books.

Investment Company Institute. 2016. *2016 Investment Company Fact Book: A Review of Trends and Activities in the U.S. Investment Company Industry*. 56th ed. www.ici.org/pdf/2016_factbook.pdf.

Jones, Jeffrey M. 2017. "U.S. Stock Ownership Down among All but Older, Higher-Income." *Gallup Economy*. (May 24): www.gallup.com/poll/211052/stock-ownership-down-among-older-higher-income.aspx.

Lucas, Robert. 1976. "Econometric Policy Evaluation: A Critique." Carnegie-Rochester Conference Series on Public Policy, vol. 1:19–46.

Lucas, Robert. 2009. "In Defense of the Dismal Science." *The Economist*. (August 6): www.economist.com/node/14165405.

Lysandrou, Photis. 2014. "Financialisation and the Limits of Circuit Theory." Paper presented at the Annual Workshop of the Post-Keynesian Economics Study Group, SOAS, University of London. (May): http://postkeynesian.net.

Mankiw, N. Gregory. 2012. *Macroeconomics*. 8th ed. Basingstoke: Worth Publishers.

Mehrling, Perry. 2015. "The Inherent Hierarchy of Money." Chapter 21 in *Social Fairness and Economics* edited by Armon Rezai, Lance Taylor, and Tom Michl. Abingdon: Routledge.

Michell, Jo. 2016. "Do Shadow Banks Create Money? 'Financialisation' and the Monetary Circuit." *Metroeconomica,* vol. 68, no. 2:354–377.

Minsky, Hyman P. 1986. *Stabilizing an Unstable Econ* omy. New Haven, CT: Yale

University Press.
Mosler, Warren. 2010. *The Seven Deadly Innocent Frauds of Economic Policy*. Christiansted, USVI: Valance Co.
Mosler, Warren. 2013. *Soft Currency Economics II*. CreateSpace Independent Publishing Platform.
Muth, John F. 1961. "Rational Expectations and the Theory of Price Movements." *Econometrica*, vol. 29, no. 3 (July):315–335.
Passarella, Marco Veronese. 2014. "Financialisation and the Monetary Circuit: A Macro-Accounting Heterodox Approach." *Review of Political Economy*, vol. 26, no. 1:128–148.
Piketty, Thomas. 2014. *Le Capital au XXI Siècle*. Paris: Editions du Seuil. English title: *Capital in the Twenty-First Century*. Cambridge, MA: Belknap Press.
Raberto, Marco, Andrea Teglio, and Silvano Cincotti. 2012. "Debt, Deleveraging and Business Cycles: An Agent-based Perspective." *Economics – The Open-Access, Open-Assessment E-Journal*, 27:50.
Realfonzo, Riccardo. 2012. "Circuit Theory." In The Elgar Companion to Post Keynesian Economics. Edited by J.E. King. Cheltenham: Edward Elgar:87–92.
Romer, David. 2013. "Preventing the Next Catastrophe: Where Do We Stand?" Presentation at the IMF conference "Rethinking Macro Policy II: First Step and Early Lessons." (May 9).
Sawyer, Malcolm C. 2013. "Endogenous Money, Circuits and Financialization." *Review of Keynesian Economics,* vol. 1, no. 2 (April):230–240.
Seccareccia, Mario. 2012 "Financialization and the Transformation of Commercial Banking: Understanding the Recent Canadian Experience before and during the International Financial Crisis." *Journal of Post Keynesian Economics*, vol. 35, no. 2 (December):277–300.
Turner, Adair. 2013. "Credit, Money and Leverage: What Wicksell, Hayek and Fisher Knew and Modern Macroeconomics Forgot." Speech made at conference "Towards a Sustainable Financial System." Stockholm School of Economics. (September 12): https://cdn.evbuc.com/eventlogos/67785745/turner.pdf .
Tymoigne, Eric. 2006. "The Minskyan System, Part III: System Dynamics Modeling of a Stock Flow-Consistent Minskyan Model." Working Paper 455. The Levy Economic Institute of the Bard College.
Werner, Richard A. 1992. "Towards a Quantity Theory of Disaggregated Credit and International Capital Flows." Paper presented at the Royal Economic Society Annual Conference, York, April 1993, and at the 5th Annual PACAP Conference on Pacific-Asian Capital Markets, Kuala Lumpur, June 1993.
Werner, Richard. 1997. "Towards a New Monetary Paradigm: A Quantity Theorem

of Disaggregated Credit, with Evidence from Japan." *Kredit und Kapital*, vol. 30, no. 2 (July):276–309.

Concluding Remarks

In this book we outlined the theory of money trying to (1) illustrate both the historical positioning of modern money as well as the present conceptual framework of money, (2) discuss the engineering problem in designing new forms of money, and (3) explore the interaction of money with the real economy.

Money developed historically through a combination of debt/credit relationships, barter, and the invention of coins. Economists tend to favor a path from barter to the coinage of money while anthropologists tend to put the accent on credit. Most likely, the two elements of barter and credit were present in ancient communities. At the time of the Roman Empire, while a sophisticated coinage system had already been developed to handle exchanges of relatively small amounts, a sophisticated banking system was developed to handle large financial transactions based on credit. After the general collapse of commerce and industry in the Middle Ages, monetary systems based on coins and credit were reinstated in the early Renaissance period and developed progressively into a system of gold-backed currencies and bank credit and finally into the present system of fiat money and bank credit.

The engineering problem of money is to create a system of tools that facilitate economic transactions. The solution that is currently adopted is to create tokens that represent the purchasing power of individual agents, either households or firms. Tokens took the form of state-issued banknotes and bank deposits. Though bank deposits are convertible into banknotes, their key function is to transfer purchasing power, thereby enabling economic transactions. The key issues are how to create and how to distribute these tokens. The current solution to creating and distributing money is based on two separate processes: commercial banks extending loans and central banks purchasing assets. Banknotes, which are printed by the central banks, represent only a small fraction of the money in circulation and can be obtained by the public in the aggregate only by debiting bank deposits. The stock of money of advanced economies is primarily formed by bank deposits obtained either from loans or from sales of assets to central banks.

Asset purchases by central banks grew enormously in the aftermath of the 2007–2008 financial crisis, as central banks implemented quantitative easing. Deposits in commercial banks due to quantitative easing are now a considerable fraction of all bank deposits.

As for lending, the landscape has also changed significantly in the last 50 years. Households, not firms, have become the major borrowers in countries including the U.S. and the U.K. as they take out long-term loans to buy homes and increasingly

short-term loans to finance consumption. The share of consumption financed by wages has become progressively smaller with respect to the share of consumption financed by corporate profits and borrowing. At the same time, the margins and profits of firms have increased, creating fast-growing subsectors of the economy. The circuit of money has changed. Firms increasingly finance their operations with cash created by households' borrowing.

Economies have become progressively more complex in terms of products and commercial relationships. Growth is increasingly due less to quantitative growth and more to innovation that produces greater complexity. In complex highly innovative economies, measuring inflation becomes increasingly difficult. Inflation, as currently measured, is a concept applicable only to the most static parts of the economy. With growth concentrated in sectors where inflation cannot be measured, with an increasing fraction of the newly created money staying in financial markets without reaching the real economy, we have witnessed low inflation, economic growth in line with historical averages, and financial markets growing at a much higher speed than the real economy. But, as we have suggested, this "average" picture is misleading. Growth has not been equally distributed. Recurrent crises signal the intrinsic instability of economic systems increasingly driven by the creation of money.

We believe that stock-flow consistent models will help in gaining a better understanding of economic dynamics in these new non-equilibrium situations. In particular, multiagent stock-flow consistent models might offer a reasonably realistic simulation of highly complex economies.