On the Origins and Evolving Role of Money

by

JEAN-PAUL CHAVAS AND DANIEL W. BROMLEY∗

“There is no denying that views on money are as difficult to describe as shifting clouds.” Joseph SCHUMPETER [1954, p. 289]

This paper explores the origins and evolving role of money. It stresses the role of money in solving coordination problems. This role is present with or without exchange. A historical overview of money stresses the evolutionary aspects of monetary institutions. The analysis develops a general evaluation of economic efficiency under transaction costs and imperfect information, and in the presence of public goods and externalities. It identifies a nonneutral role for money. It argues that the efficiency gains from exchange are only part of the benefits of a monetary economy. (JEL: E 4, E 5, G 2, N 2)

1 Introduction

Money is often seen as a means of eliminating the inefficiency of barter by reducing transaction costs (MENGER [1892], FELDMAN [1973], JONES [1976]). Other writers have offered a more complicated account (GALBRAITH [1975], DAVIES [2002], McMILLAN [2002], INGHAM [2004]). The role of money as a medium of exchange has also been investigated in the context of decentralized search and random matching (KIYOTAKI AND WRIGHT [1989], STARR (ed.) [1989]) with implications for the endogenous rise of marketplaces (HAYASHI AND MATSUI [1996], MATSUI AND SHIMIZU [2005]).¹

¹ These decentralized search models have stimulated research on the economics of money. This includes STARR (ed.) [1989]; KAHN, MCANDREWS, AND ROBERDS [2005]; and HOWITT [2005]. KAHN, MCANDREWS, AND ROBERDS [2005] investigated the role of information. They show that when the legal structure cannot provide perfect enforcement, money can be useful in that its privacy reduces the potential for opportunistic behavior. HOWITT [2005] analyzed the role of spatial separation and exchange costs in the endogenous rise of market places.

Journal of Institutional and Theoretical Economics
JITE 164 (2008), 624–651 © 2008 Mohr Siebeck – ISSN 0932-4569
This emphasis on exchange obscures the historical fact that money appeared prior to market exchange (INNES [1913], KEYNES [1930]). Babylonian clay tablets represented statements of indebtedness, and this clearly predates the first coins used in exchange (INNES [1913]). In addition, money has long played a role as a measure of value in the enforcement of rights and correlated duties. Since exchange cannot occur without the prior existence and enforcement of property rights, this suggests that money had its origin prior to the development of markets and market exchange. Indeed, there is historical evidence that money did not arise as a consequence of autonomous individual choices (e.g., SELGIN [2003], INGHAM [2004]). Finally, besides the historical account, there is some evidence that exchange may be only a weak motivator for the establishment of money. For instance, LEIBENSTEIN [1966] has argued that the gains from competitive exchange tend to be relatively small. In addition, a large part of the gains from trade liberalization have been associated with productivity improvements that go beyond the classical measures of gains from trade (e.g., KELLER [2000], PAVCNIK [2002], FERREIRA AND ROSSI [2003]).

Thus, despite a great deal of study, the original purpose of money – its foundational role in an economy – remains contested. In studying the history of money it is easy to see that money has, at various times and places, played several roles – as a measure of value, as a medium of exchange in market transactions, as a store of value, and as a means of payment. This raises the following questions: (1) is it possible to establish a theoretical argument that can illustrate the different roles that we know money has played in human history; and (2) can we then isolate parts of that theoretical structure so that distinct aspects of that structure can be brought into concordance with the known historical record?

This paper explores the linkages between money and economic efficiency. The analysis is presented under general conditions to provide new insights into the multiple functions of money, with a special focus on its nonexchange roles. While downplaying the importance of exchange may seem inappropriate in the current context of globalization, we argue that treating money simply as a means of facilitating exchange hides important insights into the determinants of economic efficiency. In particular, we stress the efficiency-enhancing role of money in solving coordination problems. Since coordination problems arise in the presence of externalities and/or public goods, special attention is given to such situations. This allows us to investigate the efficiency-enhancing role of money even in the absence of exchange. However, this is not inconsistent with the role of money in market economies. In particular, our approach is presented with allowance for frictions in transfers. We develop the arguments with a special focus on the role of money in solving coordination problems associated with externalities or public goods. In this context, we argue that money has a nonneutral role in the functioning of an economy.

The paper is organized as follows. A motivation for our analysis is presented in sections 2 and 3. Section 2 provides an overview of the historical origin of money. Section 3 presents a contemporaneous example to illustrate how money can help solve coordination problems in teams even in the absence of exchange. Section 4 develops a general model of an economy containing both private and public goods,
and exhibiting externalities. Implications for the properties of efficient allocations under alternative monetary and exchange regimes are examined in sections 5 and 6. We show that reducing transaction costs remains important as a way to improve efficiency. Under bounded rationality, we also examine the role of money in the management of coordination failures and its contributions to improving efficiency. Section 7 concludes.

2 The Historical Account

Scholars have grappled with the concept and role of money for centuries. As mentioned in the introduction, a popular view among economists is that money arose as a medium of exchange (e.g., Menger [1892], Feldman [1973], Jones [1976], Kiyotaki and Wright [1989], Starr (ed.) [1989], Hayashi and Matsui [1996], Matsui and Shimizu [2005]). By reducing transaction costs (compared to barter), money is alleged to be central to the emergence of market exchange and the full flowering of markets – both for trade among communities, and for exchange within communities. On this account, money contributes to economic growth by generating benefits from external trade. Below, we briefly review the historical origins of money under the premise that these origins are not well understood. Given the complex evolution of money through the ages (as discussed by Galbraith [1975], Davies [2002], Ingham [2004], and others), this review is highly selective.

2.1 Early Developments

Following Davies [2002, p. 28], we define money as “an institution that is widely used for making payments and accounting for debts and credits.” The Babylonian clay tablets (dating around 3000 B.C.) mentioned above acknowledged indebtedness and were thus the earliest mention of a specific form of money. These tablets provided bookkeeping for debt payments related to agricultural surplus (e.g., grain) or labor in a hierarchically organized society (Innes [1913]). This predates by more than 2000 years the appearance of coinage and its connection with exchange. In the absence of coinage, most transactions in Babylon involved the assignment of credit related to grain stocks within a command economy (Ingham [2004, p. 95]). The temple–farmer debts were personal and could not be transferred to third parties. Coins did not appear until around 640 B.C. (in Lydia), at which time the minting of coins spread rapidly to the city-states of ancient Greece. Athens started producing coins around 610 B.C. Coinage not only facilitated trade, but – because it could be easily hidden to evade local authorities – also contributed to a weakening of the bureaucratic control of the local elite. Coinage also helped to support the first large-scale instance of wage labor – mercenary soldiers (Cook [1958]). Alexander the Great (336–323 B.C.) established a stable coinage throughout the Hellenic empire.

2 True coinage was first invented around 640 B.C. in Asia Minor in the ancient state of Lydia (now western Turkey) (Davies [2002, p. 62]).
and used it to pay his soldiers. With the imposition of taxation after conquest, the defeated people would use the coins received from soldiers to pay taxes to the new central authority (INGHAM [2004, pp. 99f.]). Setting taxes payable in coins was apparently more effective than the use of forced labor.

The Romans were the next major state to establish coinage on a large scale. The first Roman coins were minted in 269 B.C. in Rome. In contrast with earlier empires, direct control was much weaker. Where administrative organization allowed sufficient coercion, the provinces were forced to pay taxes to fund the central government, and to pay soldiers (INGHAM [2004, p. 101]). In regions where taxation was less coercive, manorial relations persisted. For many centuries the balance between returns to Rome and the cost of military expansion was positive. However, this balance eventually turned negative. A growing inability to supply precious metals for coins, to collect taxes, and to pay military expenses contributed to the fall of the Roman Empire (INGHAM [2004, pp. 102–105]). After the collapse of Rome in 476 A.D., the failure of the barbarian chiefs to coerce tax payments led to the disintegration of the Roman monetary and economic space. In large areas of the former empire barter resumed. The original Roman money practically disappeared, and where it existed, it was strictly an accounting device.

2.2 Medieval Europe

The minting of coins did not resume on a large scale in Europe until the eleventh century, and it arose in conjunction with the development of kingdoms and ecclesiastical jurisdictions (BLOCH [1954]). At that time, Europe was a collection of feudal fiefdoms where long-distance overland trade was nearly nonexistent, and traders were vulnerable to attacks by brigands and extortion by feudal lords. During this period there was a multiplicity of coins, each minted by different authorities, with different sizes, weights, and silver content. While early medieval banks took cash deposits for safekeeping and provided money-changing services to traders facing a monetary anarchy of multiple coinages, they did not issue credit. The twelfth century saw the beginning of a period of monetary innovations. Interestingly, some of these innovations developed during the large-scale military campaigns of the Crusades. This is exemplified by the story of the Knights Templar, an order of warrior monks created in 1118 to keep the roads and highways safe, with a special regard for the protection of pilgrims. The Knights Templar had an exceptional degree of autonomy. Being responsible only to the Pope, the Templars were beyond the control of bishops and kings. They soon acquired considerable wealth and power – they had their own army, a fleet of ships, and storehouses, and they occupied castles at a number of strategically located ports and towns throughout western Europe (DAVIES [2002, p. 155]). The Templars bolstered the development of an extensive communication

3 For example, the Genoese lira did not have the same value as the English pound, the French livre, or the Milanese lire.

4 During this period, the Catholic Church forbade Christians to charge interest on loans. As a result, money lending was practiced mostly by Jewish entrepreneurs.
network throughout Europe and the Middle East, and they carried out merchant and banking activities supporting both commerce and the Crusades. By the end of the thirteenth century Templars had become a multinational conglomerate with interests in many commercial activities.

By controlling the trade routes of Europe they made long-distance trade possible in relative safety (and free from extortion). The exchange of knowledge and technology between the Muslim world and Europe increased, and the Templars stimulated the growth of financial centers in western Europe, especially in Italy and France. As one of the main financial intermediaries, the Knights Templar contributed to the widespread use of the modern bill of exchange among merchants and bankers (DAVIES [2002, p. 156]). This stimulated the emergence of a newly prosperous merchant class and thereby altered the balance of power away from feudal lords.

For a fee the Templars provided a range of financial services: (1) they facilitated and financed the pilgrimage trade; (2) they arranged safe transfer of funds for international and local trade; and (3) they extended credit to merchants, bishops, princes, and kings. Predictably, their economic power came to be resented by the kings. On Friday, October 13, 1307, King Philip – deeply indebted to the Templars – had all French Templars arrested and charged with many crimes. In 1312 King Philip persuaded the Pope to disband the Templars, and two years later he had the remaining French Templar leaders executed.

2.3 The Rise of Modern Banking

By the fourteenth century Mediterranean city-states had developed public banks consisting of alliances among bankers, moneychangers, and city governments. This alliance was dependent on the cohesion of the local ruling oligarchy. The city government made public loans and became the bankers’ largest clients. The government required bankers to make their records available. The situation was different in the European kingdoms, where monetary policy was controlled by the monarchs. In support of the monarch’s economic power, depreciating the value of coins was an alternative to raising taxes. This implied an uneasy balance between rulers and bankers. Monarchs saw merchants and bankers as competitors who threatened their economic control. In 1577, the French state reasserted sovereign control of its monetary system, which reduced profit opportunities for exchange bankers (INGHAM [2004, p. 120]).

Until the sixteenth century credit relations were mostly personal: they were made before a notary and witnesses. This meant that debt was not transferable (marketable). During the sixteenth century, with the appearance of transferable bills of exchange representing a promise to pay denominated in local money, depersonalized debt relations developed in the Italian city-states. However, transferability was restricted to the network of bankers. Indeed, bankers opposed the broad circulation of such bills, as it threatened their profit opportunities (INGHAM [2004, p. 120]). A more widespread use of depersonalized debt occurred in Holland in the middle of the sixteenth century, and later in England.
As suggested by Selgin [2003] and Ingham [2004], the gradual evolution of transferable bills of exchange into credit money would not have been possible without a partnership between bankers and the state. Indeed, without a wider base, transfers of bills of exchange would have remained restricted to small networks of bankers and merchants. The economic space required for widespread depersonalized exchange was eventually provided by the state. This occurred where a suitable balance of power existed between bankers and monarchs.

England was the first state where credit money was successfully established as public currency. The partnership between the English monarch and bankers arose out of the political struggle over coinage (controlled by the monarch) and credit (controlled by bankers). First, royal prerogatives implementing loose monetary policies (e.g., the debasement of coins) created financial insecurity that proved costly to bankers. Second, English kings attempted to control bankers by restricting monetary exchanges. Third, while the monarchs benefited from bankers’ loans, there was a real threat of default on their loan payments. These factors induced a compromise between the different parties, leading to the joint use of credit money and coinage, and to the foundation of the Bank of England in 1694. Out of the compromise, the transferability of bills of exchange was generalized, while the creation of the Bank of England limited the monetary control exercised by the monarchy. The integration of coin money and credit money proved to be the cornerstone of modern monetary institutions. This also set the stage for independent central banks operating under rules designed to reduce political interference (e.g., the U.S. Federal Reserve, created in 1913, or the European Central Bank, established in 1998). Finding a proper balance in monetary policy between the interests of the state and the interests of banks remains an important aspect of the efficient functioning of an economy.5

3 Solving Coordination Problems: A Contemporary Example

The role of money in solving coordination problems has been emphasized in the analysis of primitive economies (Einzig [1948], Dalton [1965], Hodges [1988]). For example, in his book Primitive and Peasant Markets, Richard Hodges writes: “The temptation in modern society, and the temptation when studying past societies, [...] is to adopt a one-dimensional approach to money. Yet, in practice, money is simultaneously an aspect of relations between persons, and an object detached from persons” (Hodges [1988, p. 96]). But, is the role of money in solving coordination problems still relevant in modern societies? We argue below that it is. This is illustrated through a contemporary example.

---

5 For example, the Great Depression of the 1930s was created in part by behavior of the U.S. Federal Reserve that was “too closely aligned” with the narrow interest of bankers (Friedman and Schwartz [1963]). At the other extreme, the cost of “excess control” by a state is illustrated by the recent economic crisis in Argentina (Ingham [2004]).
Consider the story of Coach Ken Carter. Mr. Carter was hired in 1997 to coach a group of basketball players at Richmond (California) High School, taking over from a kind and gentle predecessor who managed a record of 4–22 in his final season. Coach Carter acquired plausible talent possessed of what a history teacher would call an “attitude problem” but what even an average coach would immediately recognize as a coordination problem: individual players on the basketball team did not perform to the best of their ability. Coach Carter was no average coach. He had been a championship player in his days at Richmond, and his name adorned a banner from the rafters marking two consecutive years as top scorer. Coach Carter possessed the necessary skills to solve coordination problems.

Carter’s realm was a small closed economy. Moreover, there was no exchange inside Carter’s economy. Coach Carter solved his coordination problem by ruling that individual behavior inimical to the goals of the group (the economy) would be addressed using one debt instrument: pushups. There were 100-pushup lapses, 500-pushup lapses, 1,000-pushup lapses, 2,000-pushup lapses, and in the extreme, a member of the economy could owe Coach Carter over 2,500 pushups. Often it was the deviant agent who paid pushups to Coach Carter, but at other times all members of the economy paid pushups to Coach Carter for the lapses of a single member of the group. Coach Carter was eager to collect pushups, but not because he could accrue pushups in the bank and then use them as a medium of exchange. Nor were pushups a store of value. The sole function of pushups was to address coordination problems. When Coach Carter found that his players were not attending classes, pushups seemed inadequate and so he locked his team out of the gym, and threatened to cancel the rest of the season. Predictably, parents and the school board forced him to relent. The players – in an unexpected show of solidarity toward their Coach, and in a surprising commitment to scholarship – moved their desks into the gym and did their schoolwork there rather than practice. In his first season, Coach Carter took the Richmond team to the state tournament, where they lost by one point to the top-seeded team on a last-minute shot.

Coach Carter’s story provides perfect motivation to discuss the role of money. Taking the pushup as the unit of money, the world of Coach Carter and his players is an autarkic economy in which pushups play a profound role in solving coordination problems: modifying individual behavior so as to improve team performance. Yet pushups were just a debt instrument. Pushups imply serious sacrifice to those who perform them. For this reason, they would not be voluntarily accepted by each student. The pertinent story of Coach Carter concerns how to set up new rules of behavior within an autarkic team, supported by pushups as a payment scheme.

In this story, Coach Carter was behaving as a coordinator who set up new rules for members of the team. He created new debt structures and established new rights with respect to their attribution, enforcement, and payment. We see here innovations in property rights and contract design within the team. In this context, Coach Carter

---

6 The story was recently made into a movie: “Coach Carter,” Paramount Pictures, 2005.
designed and enforced new debt contracts involving agents in the economy. Note that this requires the existence of a hierarchy, where the coordinator is acknowledged as the person with the authority to write contracts with each of the agents. In this particular case, the terms of the contracts are simple. The coordinator decides the specific performance expected of each with whom he has a contract, and the debt payment due when there is a breach of the contract.

Why did Coach Carter demand payment in pushups? As the coordinator, he needed some unit of account to specify the debt payments (fines) associated with each agent. Coach Carter chose pushups because this was a unit of account that each player had the capacity (ability) to give. The coordination problem was to create a set of institutions (rights and duties) that could be applied to all agents, with three important characteristics: (1) each agent has the ability to pay the fines; (2) monitoring was easy (low cost); and (3) Coach Carter did not benefit privately from collecting pushups. The athletes were the owners of their labor power and thus could be made to surrender units of it to the Coordinator on demand. Labor power is a private good with use rights that are contractible. As such, it is well suited for a debt payment. In this context, the Coordinator set up new rules that refined the athletes’ use rights of their labor power, and that created a performance–debt contract that set minimal performance standards for each of them, with debt payments being triggered when those standards were not met. Using pushups as units of money, Coach Carter’s debt contracts contributed to a large improvement in the performance of the team. Importantly, this took place without individual exchange of valuable services, and without private wealth accumulation.

The story of Coach Carter illustrates that neither exchange nor value accumulation is required for money to play an important role in the functioning of an economy. Debt burdens can be assigned at will by the coordinator, and the debtor can be made to liquidate that obligation almost instantaneously. Debts are not transferable. Money is present, exchange is not.

4 Model of an Economy

Having presented historical and contemporary evidence about the multiple roles of money, we now develop a formal model to identify the efficiency implications associated with these multiple roles. In this section, we present a model of an economy composed of multiple agents involved in the allocation of private goods as well as public goods. Our model considers the case of frictions in transfer that can affect the feasibility of exchange. It also considers the case of externalities and public

---

7 Note that team performance (e.g., number of points scored in a game) is not individually contractible. Indeed, team performance depends in part on luck, and in part on synergies among the team members. Thus, team performance exhibits characteristics of public goods that cannot be easily associated with specific individuals. In other words, team performance cannot be easily appropriated.
goods. It is developed under a state-contingent approach representing uncertainty and asymmetric information. The model provides a framework to examine the role of money with and without exchange.

Consider a set $N$ of $n$ individuals $N = \{1, ..., n\}$ facing uncertainty. Following DEBREU [1959], the uncertainty is represented by $S$ mutually exclusive states of nature: $s = 1, ..., S$. The $n$ individuals face $m$ private goods and $k$ public goods. Denote the state-contingent consumption goods obtained by the $i$th individual by

$$x_i = (x_{ai}, x_b),$$

where

$$x_{ai} = (x_{ai}(1), ..., x_{ai}(S)) \in X_a \subset R^{mS}, \quad x_{ai}(s) = (x_{ai1}(s), ..., x_{aim}(s)) \in R^m,$$

is the vector of $m$ private goods obtained by the $i$th individual under state $s$, $s = 1, ..., S$. And denote the state-contingent public goods by

$$x_b = x_b((1), ..., x_b(S)) \in X_b \subset R^{kS},$$

where

$$x_b(s) = (x_{b1}(s), ..., x_{bk}(s)) \in R^k$$

is the vector of $k$ public goods obtained under state $s$, $s = 1, ..., S$. For the $i$th individual, the set $X_{ai} \times X_b$ is the state-contingent consumption possibility set for the bundle $x_i = (x_{ai}, x_b)$. Thus, consumption goods in the economy are given by

$$x \equiv (x_{ai}, x_b) \in X,$$

where

$$x_{ai} \equiv (x_{ai1}, ..., x_{aim}) \quad \text{and} \quad X \equiv X_{ai1} \times ... \times X_{aim} \times X_b.$$

The consumption goods $x \in X$ must be produced. We consider the case of a production economy supported by a set $J$ of $q$ firms, $J = \{1, ..., q\}$. Denote by $y_{aj} = (y_{aj}(1), ..., y_{aj}(S))$ the state-contingent private goods produced by the $j$th firm, where $y_{aj}(s) \in R^m$ is the $j$th firm’s production of private goods under state $s$, $s = 1, ..., S$. And denote by $y_b = (y_{b1}(1), ..., y_{bq}(S)) \in R^{kS}$ the vector of state-contingent public goods produced, where $y_b(s)$ is the quantity of $k$ public goods produced under state $s$, $s = 1, ..., S$. Then, if we let $y_{aj} = (y_{aj1}, ..., y_{ajq})$, the pair $y \equiv (y_{ai}, y_b)$ represents the quantity of all goods produced in the economy.

To represent exchange, it will be convenient to treat individuals and firms as economic agents. Denote by $t_{ij} = (t_{ij}(1), ..., t_{ij}(S)) \in R^{nS}$ the vector of state-contingent private goods exchanged from the $i$th agent to the $j$th agent, with $i, j \in N \cup J$, where $t_{ij}(s) \in R^{mS}$ is the quantity of private goods transferred from agent $i$ to $j$ under state $s$, $s = 1, ..., S$. Then

$$t \equiv \{t_{ij} \in R^{nS}; i, j \in N \cup J\}$$

represents all exchanges of private goods among all agents. In the case where transfers require the use of resources, let $z_j = (z_j(1), ..., z_j(S)) \in R^{mS}$ denote the
vector of state-contingent private goods used by the \( j \)th agent to accomplish the transfers, \( j \in N \cup J \), where \( z_j(s) \in \mathbb{R}^m \) is the quantity of private goods used by the \( j \)th agent under state \( s \), \( s = 1, \ldots, S \). Then
\[
z = \{z_j \in \mathbb{R}^m : j \in N \cup J\}
\]
represents the private goods used to support exchange.

The state-contingent approach makes it explicit that decisions depend on the states of nature. Uncertainty arises when some states of nature are not known at the time a specific decision is made. This available information can be represented by partitions of the state space \( \{1, \ldots, S\} \). At one extreme, the greatest amount of information corresponds to the finest partition \( P^+ \), which has \( S \) elements. At the other extreme, the least amount of information corresponds to the coarsest partition \( P^- \), which has one element. In intermediate situations, the number of elements of a partition of \( \{1, \ldots, S\} \) can vary between 1 and \( S \).

The allocation in the economy involves the vector \( w \equiv (x, y, z, t) \), where \( x = (x_a, x_b) \) are consumption goods, \( y \equiv (y_a, y_b) \) are production goods, \( z \) are the resources used in the exchange process, and \( t \) denotes exchange. The allocation decisions can be written as
\[
w = \{w(s) : s = 1, \ldots, S\},
\]
where
\[
w(s) = (w_1(s), w_2(s), \ldots)
\]
and \( w_r(s) \) represents the \( r \)th decision made under state \( s \), \( s = 1, \ldots, S \). Suppose that the \( r \)th decision is made under an information set represented by the partition \( P_r \) of \( \{1, \ldots, S\} \). Situations where two different states \( s \) and \( s' \) are in the same element of the partition \( P_r \) are ones where the decision-maker choosing \( w_r \) cannot distinguish between \( s \) and \( s' \). Alternatively, finding that \( s \) and \( s' \) are in different elements of the partition \( P_r \) means that the decision-maker choosing \( w_r \) can distinguish between \( s \) and \( s' \). Denote the information structure supporting all decisions by \( \mathcal{G} = \{P_r : r = 1, 2, \ldots\} \). This allows for situations of asymmetric information in which different decision-makers have access to different information.\(^8\) And in a dynamic context (where the decisions are made over time), this allows for information-gathering activities and learning (e.g., learning by doing).

For a given information structure \( \mathcal{G} \), the state-contingent technology supporting the economy is represented by the feasible set \( T(\mathcal{G}) \), where \((y, z, t) \in T(\mathcal{G}) \) means that \((y, z, t) \) are feasible under information \( \mathcal{G} \). Note the generality of the technology. It allows for private goods as well as public goods. As \( T(\mathcal{G}) \) represents a joint production–distribution process among all firms, it allows for external effects across firms. It allows for positive transaction costs, where exchange among agents may require the use of resources. Finally, it provides a state-contingent representation of

\(^8\) Thus, our model allows for both individual heterogeneity and asymmetric information. Incorporating such characteristics in the analysis of monetary economies has been seen as an important part of improving current models of monetary economies (e.g., Frydman and Goldberg [2007], Mishkin [2007]).
the uncertainty facing all agents in the economy. Throughout, we focus our attention on information structures $\wp$ that satisfy $T(\wp) \neq \emptyset$. We also assume that the set $T(\wp)$ is closed.

For simplicity, we consider the case where only firms produce (implying that $y_{ai} = 0$ for $i \in N$) and where only individuals consume (implying that $x_{ai} = 0$ for $j \in J$). In this context, under an information structure $\wp$, a feasible allocation $w \equiv (x, y, z, t)$ satisfies $x \in X$, $(y, z, t) \in T(\wp)$,

\begin{align}
\sum_{j \in N \cup J} t_{ij} &\leq y_{ai} - z_i, \\
x_{ai} &\leq \sum_{j \in N \cup J} t_{ji} \\
\end{align}

for $i \in N \cup J$,

\begin{align}
x_{ai} &= y_{ai}, \\
\end{align}

and

\begin{align}
w_r(s) &= w_r(s'),
\end{align}

when $s$ and $s'$ ($s \neq s'$) are in the same element of $P_r$, $r = 1, 2, \ldots$.

The inequalities in (1a)-(1b) represent the balance of transfers in private goods. That is, $\sum_{j \in N \cup J} t_{ij} \leq y_{ai} - z_i$ in (1a) requires that the transfer outflows from the $i$th agent cannot be greater than its production, net of the resources used to accomplish the transfers. And $x_{ai} \leq \sum_{j \in N \cup J} t_{ji}$ in (1b) guarantees that the $j$th individual’s consumption of private goods cannot be greater than his/her total transfer inflows. Equation (1c) implies that the public goods $x_0$ must be produced. Note that equations (1a)-(1c) exhibit three important characteristics. First, they allow for frictions in transfers (represented by $z$). Second, they allow for private as well as public goods under uncertainty and asymmetric information. Third, they include as special cases situations of autarky when there is no exchange among firms (i.e., when $t_{ij} = 0$ for $i \in J, j \in J, i \neq j$). Think of this as a group of individuals supplying labor to productive activities that generate goods consumed by members of the group (e.g., a primitive economy; or the case of subsistence farms in developing countries, with limited access to markets). It can also represent the environment managed by Coach Carter.

Equation (2) imposes information restrictions. It states that each decision $w_r$ can depend only on the states that can be distinguished by the decision-maker (e.g., RADNER [1968]). At one extreme, if $w_r$ is chosen under perfect information, then $P_r = P^+$ and $w_r$ is chosen ex post (as equation (2) becomes vacuous, allowing $w_r$ to vary freely across states). At the other extreme, under no information, $P_r = P^-$ would have a single element, and $w_r$ would be chosen ex ante (as equation (2) would restrict it to be the same under all states). In intermediate situations, $P_r$ has more than one element but less than $S$ elements, reflecting situations of partial information. We focus attention on situations of active learning where information is costly. This is represented by $T(\wp') \subset T(\wp)$ for any $\wp'$ that is at least as fine as $\wp$. Then,
On the Origins (2008) 635

Let the preferences of the $i$th individual be represented by the utility function $U_i(x_{ai}, x_b), i \in N$. We assume that the utility function $U_i(x_{ai}, x_b)$ is continuous and quasiconcave in $(x_{ai}, x_b), i \in N$. Consider a reference bundle of private goods $g \in R_{g}^{n}$ satisfying $g \geq 0$ and $g \neq 0$. Following Luenberger [1992], define the $i$th individual benefit function

$$B_i(x_{ai}, x_b, u_i)$$

where $u_i$ is a reference utility level for the $i$th individual, $i \in N$. The benefit function $B_i(x_{ai}, x_b, u_i)$ in (3) measures the number of units of the reference bundle $g$ that the $i$th consumer is willing to give up to obtain the consumption goods $(x_{ai}, x_b)$ starting from utility level $u_i, i \in N$. Given a constant reference bundle $g$, benefits can be added across individuals. This generates the aggregate benefit

$$\sum_{i \in N} B_i(x_{ai}, x_b, u_i),$$

which measures the total number of units of $g$ the $n$ individuals are willing to give up to obtain $(x_{ai}, x_b)$ starting from utility levels $u = \{u_i : i \in N\}$. Under the quasiconcavity of $U_i(x_{ai}, x_b)$, the benefit function $B_i(x_{ai}, x_b, u_i)$ is concave in $(x_{ai}, x_b)$ (LUENBERGER [1992]). Following ALLAIIS [1943], the aggregate benefit

$$\sum_{i \in N} B_i(x_{ai}, x_b, u_i)$$

measures the economic surplus that can be redistributed among the $n$ individuals.

We say that the $i$th individual is nonsatiated in $g$ if $U_i(x_{ai} + \beta g, x_b) > U_i(x_{ai}, x_b)$ for any $(x_{ai}, x_b) \in X_i$ and any $\beta > 0$. In this context, we make the following assumptions.

**ASSUMPTION 1** For a given information structure $\varphi$, and for any feasible allocation $(x, y, z, \tau)$ satisfying $\sum_{i \in N} B_i(x_{ai}, x_b, u_i) > 0$, there is another feasible allocation $(x', y', z', \tau')$ such that

(a) $x'_i = x_{ai} + b_i g$ for all $i \in N$, and $x'_b = x_b$, where $(b_1, \ldots, b_n) \in R_{b}^{n}$ and $\sum_{i \in N} b_i = \sum_{i \in N} B_i(x_{ai}, x_b, u_i)$.

(b) the individuals for whom $b_i > 0$ are nonsatiated in $g$.

**ASSUMPTION 2** The consumption bundle $x_{ai}$ is in the $g$-interior of $X_{ai}$ for each individual, i.e., it satisfies $x_{ai} - \beta g \in X_{ai}$ for some $\beta > 0$ and for all $i \in N$.

Assumption 1 states that any positive surplus can be feasibly redistributed among the $n$ individuals. This is important in situations where transfers across individuals require the use of resources (which may imply that some redistribution schemes

---

9 In a state-contingent framework, the quasiconcavity of $U_i(x_{ai}, x_b), i \in N$, corresponds to risk aversion (DEBREU [1959]).

10 Note that the reference bundle $g$ should not be interpreted as “money.” As shown below, it will be used simply as a convenient basis for assessing economic efficiency.
are infeasible). Assumption 1 also assumes that the individuals benefiting from a surplus redistribution are nonsatiated in the reference bundle $g$. Assumption 2 limits the analysis to allocations where the $n$ individuals are not totally destitute. This assumption guarantees that each individual has a minimal access to resources, so that giving up a positive amount of the reference bundle $g$ is feasible. Below, we analyze feasible allocations $(x, y, z, t, \varphi)$ where Assumptions 1–2 hold, and where the set $X \cap \{ (y, z, t) \in T(\varphi), \text{ equations (1) – (2)} \}$ has a nonempty interior.

The first question we want to address is: what are the optimal decisions for $(x, y, z, t)$ and $\varphi$? Relying on the classical Pareto criterion, an allocation $(x, y, z, t, \varphi)$ is said to be Pareto-efficient if there does not exist any other feasible allocation that can make one individual better off without making any one else worse off. Intuitively, Pareto-efficient decisions are expected to be related to the maximization of aggregate benefit $\sum_{i \in N} B_i(x_{a_i}, x_b, u_i)$. In this context, following LUENBERGER [1995], define a \textit{maximal} allocation as an allocation $(x, y, z, t, \varphi)$ that satisfies

$$V(u) = \max_{x \in X, (y, z, t) \in T(\varphi), \text{ equations (1)–(2)}} \left\{ \sum_{i \in N} B_i(x_{a_i}, x_b, u_i) : x \in X, \right. \left. (y, z, t) \in T(\varphi), \text{ equations (1)–(2)} \right\},$$

where $u = \{u_i : i \in N\}$ and $V(u)$ is the largest possible distributable surplus. And define a \textit{zero-maximal} allocation as a maximal allocation where the utility levels $u = \{u_i : i \in N\}$ satisfy $V(u) = 0$. Following ALLAIS [1943] and LUENBERGER [1995], a maximal allocation involves the maximization of the aggregate benefit $\sum_{i \in N} B_i(x_{a_i}, x_b, u_i)$, and a zero-maximal allocation means that the maximized surplus is entirely redistributed among all consumers.

We have the following results (see the proofs in the Appendix).

**Proposition 1** Assume that Assumption 1 holds. Then, if an allocation is Pareto efficient, that allocation is zero maximal.

**Proposition 2** If an allocation is zero-maximal, then it is Pareto-efficient compared to all feasible allocations where $x_{a_i}$ is in the $g$-interior of $X_{a_i}$, $i \in N$ (i.e., where Assumption 2 holds).

Propositions 1 and 2 provide strong linkages between Pareto efficiency and zero maximality. Exploiting these linkages will prove useful for several reasons. First, these linkages are intuitive and convenient: maximizing aggregate benefit and then redistributing the resulting surplus provides a simple characterization of economic efficiency. Second, the surplus function $V(u)$ defined in (4) is a useful measure of aggregate welfare (LUENBERGER [1995], [1996]). We will make extensive use of it in our analysis below. Finally, note that Propositions 1 and 2 apply under very general conditions: they do not require market exchange; they do not assume well-defined property rights; they allow for externalities and general technology; they do not require convexity assumptions; they allow for uncertainty and asym-
metric information; and they allow for transaction costs. Thus, zero maximality provides a useful framework to investigate the efficiency of property rights and “money.”

5 Efficiency with or without Exchange or Money

Propositions 1 and 2 apply under general conditions. They show that zero maximality provides a useful characterization of Pareto efficiency in a generic economy. Below, we examine the specific role played by exchange and monetary institutions. For our purpose, let $E$ denote the exchange and monetary institutions in the economy. Such institutions are selected from among a set $\Omega$ of feasible institutions, $E \in \Omega$. For example, in primitive economies, $E$ would characterize a situation where money is absent and any exchange involves barter. In the mid twentieth century, under the Bretton Woods monetary system, $E$ would represent institutions where different countries have different currencies (managed by central banks) facing a fixed exchange rate to the dollar, with the dollar being convertible into gold at $35 an ounce. After the collapse of the Bretton Woods monetary regime in 1971, $E$ would represent fiat currencies with floating exchange rates. In each case, exchange and monetary institutions $E$ are expected to affect the functioning of the economy. Implications for economic efficiency are explored next.

Under exchange and monetary institutions $E \in \Omega$, assume that the feasible set $T(\wp)$ takes the form

$$T(\wp, E) = Y(\wp, E) \times Z(\wp, E),$$

where $Y(\wp, E)$ is the feasible set for production activities $y$, and $Z(\wp, E)$ is the feasible set for exchange activities $(z, t)$. While this imposes the restriction that production activities and trade activities are nonjoint, this assumption will simplify our analysis. It will allow us to focus our attention on two effects of the exchange and monetary institutions $E \in \Omega$. The first is through $Z(\wp, E)$, capturing the classical argument that money reduces transaction costs (e.g., Menger [1892], Feldman [1973], Jones [1976]). For example, when $E_1$ represents a barter economy and $E_2$ a monetary economy, $Z(\wp, E_1) \subset Z(\wp, E_2)$ implies that money expands the feasible set for exchange and helps stimulate trade. The second effect is $Y(\wp, E)$. It captures the influence of exchange and monetary institutions $E$ on production activities. An example is the role of money in the functioning of teams (as illustrated in our story of Coach Carter). If a switch from $E_3$ to $E_4$ helps solve coordination problems facing individuals, then $Y(\wp, E_3) \subset Y(\wp, E_4)$, reflecting that a change in monetary institutions can contribute to improved productivity.

Consider a maximal allocation given in (4) under exchange and monetary institutions $E$. Expand (4) to include the maximization of aggregate benefit with respect to $E \in \Omega$. The associated maximized surplus is $V(u)$. First, note that utilities

---

11 Importantly, Propositions 1 and 2 allow for nonconvexities in exchange activities (e.g., due to “fixed costs” in trade as analyzed by Howitt [2005]).
$u = \{u_i : i \in N\}$ satisfying $V(u) < 0$ are not attainable by all $n$ consumers. Thus, attainability implies choosing utilities $u$ such that $V(u) \geq 0$. Second, finding $V(u) > 0$ implies an inefficient allocation. It corresponds to an allocation that is necessarily inside the Pareto utility frontier. This follows from Assumption 1, which guarantees that a positive surplus, $V(u) > 0$, can be redistributed to make at least one individual better off. Third, $V(u) = 0$ identifies allocations that are on the Pareto utility frontier, with $\{u : V(u) = 0\}$ giving the largest possible utilities $u = \{u_i : i \in N\}$ that can be achieved under technology $T$ (e.g., SAMUELSON [1956]). This provides a basis for distinguishing between efficient and inefficient allocations.\footnote{As is well known, the Pareto efficiency criterion does not provide guidance to choose among efficient points on the Pareto utility frontier. Indeed, different points on the Pareto utility frontier correspond to different distributions of welfare among the $n$ individuals. Choosing a point on the utility frontier requires going beyond efficiency: it requires an evaluation of welfare distribution.}

To evaluate economic welfare, start from a feasible allocation $(x, y, z, t, g)$ under monetary institutions $E \in \Omega$. Consider the aggregate welfare measure

\begin{equation}
W(x, u, E) = V(u) - \sum_{i \in N} B_i(x_{ai}, x_b, u_i).
\end{equation}

For a given $u = \{u_i : i \in N\}$, the measure $W(x, u, E)$ in (5) measures the change in surplus associated with a change from the feasible point $(x, y, z, t, g, E)$ to a maximal allocation. By the definition (4), it follows that $W(x, u, E) \geq 0$. Thus, $W(x, u, E) \geq 0$ can be interpreted as a measure of aggregate efficiency gains: it is the number of units of the reference bundle $g$ that can be obtained by moving from the feasible point $(x, y, z, t, g, E)$ to a more efficient allocation, conditional on $u$. While there are many possible choices for $u = \{u_i : i \in N\}$, two choices appear attractive (LUENBERGER [1996]): (1) the case of \textit{compensating variations} where $u_i = U_i(x_{ai}, x_b)$, $i \in N$, i.e., where the reference utilities $u$ correspond to the initial allocation $(x, y)$; and (2) the case of \textit{equivalent variations} where $u \in \{u' : V(u') = 0\}$, i.e., where the reference utilities $u$ correspond to the subsequent zero-maximal (and thus efficient) allocation.

Three possible scenarios can arise, depending on the sign of the welfare change $W(x, u, E)$ in (5). In the first scenario, $W(x, u, E) > 0$, corresponding to situations where the change improves the aggregate benefit. Note that this is not necessarily a Pareto-improving move. However, it would be a Pareto-improving move if $B_i(x_{ai}, x_b, u_i) \leq 0$ for all $i \in N$, i.e., if no individual were made worse off by the change. Alternatively, it would not be a Pareto-improving move if $B_i(x_{ai}, x_b, u_i) > 0$ for some $i \in N$, i.e., if the change made at least one individual worse off. In the second scenario, $W(x, u, E) = 0$, corresponding to situations where the change involves a redistribution of welfare. In this context, the $i$th individual gains (loses) from the change when $B_i(x_{ai}, x_b, u_i) < 0 (> 0)$. Note that, in the presence of frictions in transfers, redistribution that requires transfers is \textit{not} efficiency-neutral. Indeed, when $W(x, u, E) = 0$, a subsequent move from a maximal allocation back to $(x, y, z, t, g, E)$ would use resources in transfers. Conditional on $u$, this would necessarily reduce the aggregate benefit. It follows that frictions in transfers imply
the existence of trade-off between efficiency and distribution. In the third scenario, \( W(x, u, E) < 0 \) corresponds to situations of aggregate efficiency loss where a Pareto-improving move is not feasible.

In general, all three scenarios are possible. Under microeconomic rationality, the only requirement is that at least one individual must benefit from the change (with \( B_i(x_{ai}, x_b, u_i) < 0 \) for some \( i \in N \)). If the individual who gains from the change is the one making the decision to switch from \((x, y, z, t, g, E)\) to a maximal allocation, then the move is individually rational. If this is associated with \( W(x, u, E) < 0 \) (scenario 3), then the gainer exhibits unproductive rent-seeking behavior: his/her personal gains are at the expense of other individuals and contribute to a decline in aggregate efficiency. Indeed, if they had a choice, the losers (with \( B_i(x_{ai}, x_b, u_i) > 0 \)) would not agree to the change. This means that rent-seeking behavior is expected to arise in situations of unequal power, where the gainers are able to impose their decisions on the losers. Alternatively, if there is at least one gainer and \( W(x, u, E) > 0 \) (scenario 1), the move will be individually rational if it satisfies \( B_i(x_{ai}, x_b, u_i) \leq 0 \) for all \( i \in N \). This corresponds to a Pareto-improving move where the benefits of the change are distributed so that no individual is made worse off. Then, microeconomic rationality is consistent with aggregate efficiency. However, in all other cases, there will be some losers (satisfying \( B_i(x_{ai}, x_b, u_i) > 0 \)) who would prefer not to see the change. This means that the change from \((x, y, z, t, g, E)\) to a maximal allocation will have to be imposed, again implying the exercise of power on the part of gainers at the expense of losers.

In the case where the point \((x, y, z, t, g, E)\) is zero-maximal, then \( W(x, u, E) = 0 \) in (5). However, when the exchange and monetary institutions \( E \in \Omega \) are not zero-maximal, under feasibility, equation (5) shows that \( W(x, u, E) > 0 \). Then changes in monetary institutions can contribute to efficiency improvements. The associated welfare gains can have two sources: the effects of \( E \) on the transaction technology \( Z(g, E) \) and/or its effects on the production technology \( Y(g, E) \). To examine the importance of each source, we consider three special cases. The first case corresponds to a situation where the monetary institutions \( E \) have no effect on the production technology, with \( Y(g) \) satisfying \( Y(g) \equiv \cap_{E \in \Omega} Y(g, E) \). Denote the associated surplus in (5) (including a maximal \( E \in \Omega \)) by \( V_Y(u) \). The second case corresponds to a situation where the monetary institutions \( E \) have no effect on the transaction technology, with \( Z(g) \) satisfying \( Z(g) \equiv \cap_{E \in \Omega} Z(g, E) \). Denote the associated surplus in (5) (including a maximal \( E \in \Omega \)) by \( V_Z(u) \). Finally, we consider the joint case where monetary institutions \( E \) have no effect at all on the economy, with \( Y(g) \equiv \cap_{E \in \Omega} Y(g, E) \) and \( Z(g) \equiv \cap_{E \in \Omega} Z(g, E) \). Denote the associated surplus in (5) (including a maximal \( E \in \Omega \)) by \( V_{YZ}(u) \). Given \( Y(g) \subset Y(g, E) \) and

---

13 This has two implications: (1) efficiency issues cannot be analyzed separately from distribution issues; and (2) if any attempt at redistribution tends to decrease efficiency (due to the resources used in transfers), then moving away from a status quo distribution is more difficult than typically assumed.

14 This offers a definition of “power” as the ability of an agent to impose unwanted welfare losses on others.
$Z(\wp) \subset Z(\wp, E)$ for all $E \in \Omega$, it follows from (5) that $V(u) \geq V_T(u) \geq V_{TZ}(u)$, and $V(u) \geq V_Z(u) \geq V_{TZ}(u)$. This suggests the following measure and decomposition of efficiency gains from monetary institutions.

**PROPOSITION 3** The efficiency gains obtained from the exchange and monetary institutions are

\[
\begin{align*}
(6a) & \quad W^E(u) = V(u) - V_{TZ}(u) \geq 0 \\
(6b) & \quad = W_T(u) + W_Z(u) \geq 0 \\
(6c) & \quad = W'(u) + W_Z'(u) \geq 0,
\end{align*}
\]

where

\[
\begin{align*}
(7a) & \quad W_T(u) = V(u) - V_T(u) \geq 0, \\
(7b) & \quad W_Z(u) = V_T(u) - V_{TZ}(u) \geq 0, \\
(7c) & \quad W_Z'(u) = V(u) - V_Z(u) \geq 0, \\
(7d) & \quad W'_T(u) = V_Z(u) - V_{TZ}(u) \geq 0.
\end{align*}
\]

$W^E(u)$ in equation (6a) is a measure of the efficiency gain obtained from the functioning of monetary institutions. It is defined as the difference between the maximized surplus $V(u)$ and the surplus $V_{TZ}(u)$ obtained under a scenario where monetary institutions $E$ make no contribution to aggregate efficiency. In addition, equations (6b) and (6c) provide two alternative decompositions of this welfare gain. These decompositions reflect the relative welfare contribution of monetary institutions to production versus exchange activities. In (6b), $W^E(u)$ is decomposed into $W_T(u)$ and $W_Z(u)$, with $W_T(u)$ in (7a) measuring the added surplus obtained from the effects of $E$ on production activities, and $W_Z(u)$ in (7b) measuring the remaining exchange effects. Alternatively, in (6c), $W^E(u)$ is decomposed into $W'_Z(u)$ and $W'_E(u)$, with $W'_Z(u)$ in (7c) measuring the added surplus obtained from the effects of $E$ on exchange activities, and $W'_E(u)$ in (7d) measuring the remaining production effects.

The decompositions given in (6b) and (6c) raise the question: what is the relative importance of exchange activities versus production activities in the evaluation of the welfare effects of monetary institutions? The classical focus of money as exchange facilitator corresponds to the special case where $W^E(u) = W_Z(u) = W'_E(u)$. Under this scenario, monetary institutions play no role in the efficiency of production activities: $W_T(u) = W'_T(u) = 0$. Yet, as mentioned in the introduction, there is evidence that the gains from competitive exchange tend to be relatively small (Leibenstein [1966]). And recent studies have documented that a large part of economic gains from trade liberalization come from productivity improvements that go beyond the classical measures of gains from trade (see Keller [2000], Pavcnik [2002], Ferreira and Rossi [2003]). This suggests that $W_T(u)$ is more important than $W_Z(u)$ in (6b) (or alternatively that $W'_E(u)$ is more important than $W'_Z(u)$ in (6c)). This is also illustrated by the adverse effects of recent financial crises on economic performance (e.g., East Asia in 1997–1998, or Argentina in 2001–2002). For example, Ingham [2004] and Mishkin [2006] have documented
how a coordination failure between the state and the banking sector plunged Argentina in 2002 into the worst depression in its history. This stresses the need to explore the role of money within a broader economic perspective.

6 Efficiency in the Presence of Coordination Problems

We have seen that an efficient allocation can be characterized by a zero-maximal allocation, i.e., by achieving a maximal allocation (as given in equation (4)), followed by a redistribution of any remaining surplus. In this section, we explore the properties of a maximal allocation. Equation (4), being a constrained optimization problem, can be analyzed using the Lagrangian

\[
L = \sum_{i \in N} B_i(x_{ai}, x_b, u_i) + \sum_{i \in N \cup J} p^f_{ai} \left[ y_{ai} - z_i - \sum_{j \in N \cup J} t_{ij} \right] \\
+ \sum_{i \in N \cup J} p^c_{ai} \left[ \sum_{j \in N \cup J} t_{ji} - x_{ai} \right] + p_b \cdot [y_b + x_b],
\]

(8)

where \( p^f_{ai}, p^c_{ai}, \) and \( p_b \) are Lagrange multipliers associated with the constraint in (1a), (1b), and (1c), respectively. At the optimum, the Lagrange multipliers have the standard interpretation of measuring the shadow values of the corresponding constraints. Under a maximal allocation, the shadow values \( p^f_{ai} \) and \( p_b \) measure the additional aggregate benefit that can be obtained from one more unit of \( y_{ai} \) and \( y_b \), respectively, expressed as the number of units of the reference bundle \( g \). Similarly, the shadow values \( p^c_{ai} \) measure the additional aggregate benefit generated by one more unit of \( x_{ai} \).

For a given \( g \), the maximization of the Lagrangian in (8) with respect to \( x_{ai} \) and \( x_b \) yields

\[
\max_{x_{ai}} \left\{ B_i(x_{ai}, x_b, u_i) - p^f_{ai} \cdot x_{ai} : x_{ai} \in X_{ai}, \text{ equation (2)} \right\}, \quad i \in N,
\]

(9a) and

\[
\max_{x_b} \left\{ \sum_{i \in N} B_i(x_{ai}, x_b, u_i) - p_b \cdot x_b : x_b \in X_b, \text{ equation (2)} \right\}.
\]

Equation (9a) implies that \( x_{ai} \) should be chosen so as to maximize the \( i \)th consumer benefit net of cost, \( i \in N \). Assuming differentiability and interior solution, the associated first-order condition gives the familiar characterization for the optimality of private goods: for each individual \( i \), the marginal benefit \( \partial B_i(x_{ai}, x_b, u_i) / \partial x_{ai} \) must equal the corresponding shadow value \( p^f_{ai}, i \in N \). And equation (9b) implies that \( x_b \) should be chosen so as to maximize aggregate benefit net of cost. Again, assuming

\[15\] In 2002, Argentina saw its poverty rate increase to almost 50% (MISHKIN [2006, p. 123]). This makes it a depression even worse than the one the U.S. experienced in the 1930s.
differentiability and interior solution, the associated first-order condition gives the well-known condition for the optimality of public goods: the sum of marginal benefits across all individuals \( \sum_{i \in N} \frac{\partial B_i(x_a, x_b, u_i)}{\partial x_b} \) must equal the shadow value \( p_b \) (e.g., Samuelson [1954]). Equation (9a) implies that, for a given \( p \), the consumption decisions for private goods \( x_a \) can be decentralized at the individual level. However, the consumption decisions for the public goods \( x_b \) cannot be easily decentralized. Indeed, from (9b), choosing \( x_b \) requires information about the marginal benefit of all consumers affected by the public goods \( x_b \). Finally, these results apply in general, with or without money, and with or without exchange.

When the production set \( Y(\varphi, E) \) is convex, the maximal choice for \( y \) corresponds to maximizing the Lagrangian \( L \) in (8) (Takayama [1985, p. 75]). This gives the following result.

**Proposition 4** When the set \( Y(\varphi, E) \) is convex and for a given \((p, \varphi, E)\), efficient production decisions \( y \) satisfy

\[
\pi_y(p_a^f, p_b, \varphi, E) = \max_y \left\{ p_a^f \cdot y_a + p_b \cdot y_b : y \in Y(\varphi, E) \right\}.
\]

Equation (10) states that in a maximal allocation, production decisions are made in a way consistent with maximizing the aggregate value of production activities, conditional on the shadow value \( p \) of the goods \( y = (y_a, y_b) \) and on the monetary institutions \( E \). Equation (10) also defines the largest possible aggregate value of production activities. Since the shadow value \( p \) is expressed as the number of units of the reference bundle \( g \), it follows that the objective function in (10) is measured in units of value of \( g \). Note that equation (10) applies to private goods \( (y_a) \) as well as public goods \( (y_b) \). It also applies in the presence of frictions in transfers. Finally, equation (10) applies in the presence of externalities across firms. However, in general, equation (10) does not imply decentralized value maximization for each firm. Indeed, in the presence of public goods or externalities across firms, decentralized value maximization will fail to be consistent with (10). This means that public goods or externalities require coordination among economic agents (e.g., as in the case of Coach Carter discussed in section 3). In this context, equation (10) is a general form of the Coase theorem (Coase [1960]). It states that, even in the presence of public goods and externalities, efficient production decisions are made in a way consistent with maximizing the aggregate value of production activities, conditional on the shadow value \( p \) and on monetary institutions \( E \). This result remains valid under risk, and in the presence of frictions in transfers. Importantly, Proposition 4 identifies how monetary and exchange institutions \( E \) can affect production efficiency. To illustrate, compare two monetary institutions \( E \in \Omega \) and \( E' \in \Omega \) satisfying \( Y(\varphi, E) \subset T(\varphi, E') \). Then, equation (10) implies that \( \pi_y(p_a^f, p_b, \varphi, E) \geq \pi_y(p_a^f, p_b, \varphi, E') \), i.e., that institutions \( E' \) are more

\[\text{In general, the sets } Y(\varphi, E) \text{ and } Z(\varphi, E) \text{ being convex guarantees that competitive markets can support an efficient allocation (Debreu [1959], Guesnerie [1975]).}\]

\[\text{Note the importance of the convexity of the production set } Y(\varphi, E). \text{ Without it, equation (10) may not hold under a maximal allocation.}\]
effective (than E) in increasing the aggregate value of production. This shows how monetary institutions contribute to improved production efficiency when they facilitate coordination.

6.1 Efficiency in the Presence of Frictions in Transfers

Assume that the transaction set \( Z(\wp, E) \) is convex. Then, using the Lagrangian in (8), a maximal choice for \((z, t)\) must satisfy

\[
\pi_Z(p^e, p^f, \wp, E) = \max_{z, t} \left\{ \sum_{j \in N \cup J} p^e_{ai} \cdot \left[ \sum_{j \in N \cup J} t_{ji} \right] + \sum_{j \in N \cup J} p^f_{ai} \cdot \left[ \sum_{j \in N \cup J} t_{ji} \right] : (z, t) \in Z(\wp, E), \text{ equation (2)} \right\}
\]

\[
= \max_t \left\{ \sum_{j \in N \cup J} p^e_{ai} \cdot \left[ \sum_{j \in N \cup J} t_{ji} \right] + \sum_{j \in N \cup J} p^f_{ai} \cdot \left[ \sum_{j \in N \cup J} t_{ji} \right] - C(p^e_{ai}, p^f_{ai}, t, \wp, E) : \text{equation (2)} \right\},
\]

where

\[
(11b) \quad C(p^e_{ai}, p^f_{ai}, t, \wp, E) = \min_z \left\{ \sum_{j \in N \cup J} p^e_{ai} \cdot z_{ij} : (z, t) \in Z(\wp, E), \text{ equation (2)} \right\}.
\]

This generates the following result.

**Proposition 5** Under the convexity of \( Z(\wp, E) \) and given \( p_a \), the efficient choice of \( z \) should be made so as to minimize the aggregate transfer cost as stated in equation (11b), where \( C(p^e_{ai}, p^f_{ai}, t, \wp, E) \) is the transfer cost function.

Proposition 5 shows that minimizing transfer costs is an integral part of an efficient allocation. This is consistent with the analysis presented by FOLEY [1970] and HOWITT [2005]. In a monetary economy, this supports the standard argument that money plays an efficiency-enhancing role when it reduces transfer costs. Importantly, Proposition 5 identifies how monetary and exchange institutions can affect transaction costs. To illustrate, compare two monetary institutions \( E \in \Omega \) and \( E' \in \Omega \) satisfying \( Z(\wp, E) \subset Z(\wp, E') \). Then, equation (11b) implies that \( C(p^e_{ai}, p^f_{ai}, t, \wp, E') \leq C(p^e_{ai}, p^f_{ai}, t, \wp, E) \). This shows that institutions \( E' \) are more effective (than \( E \)) in reducing transfer costs. However, equation (11b) does not equate minimizing transfer costs with eliminating transfer costs. Indeed, transfers typically involve some transaction costs that cannot be totally eliminated. Proposition 5 shows that these costs depend on the monetary institutions \( E \in \Omega \) (e.g., whether money is commodity
Such institutions affect the transfer technology $Z(\varphi, E)$. This has two important implications. First, discussing the role of money (e.g., fiat money) without reference to the transfer technology is incomplete. Second, the analysis of monetary innovations must be placed in the context of innovations in the transfer technology (e.g., reductions in information cost). Thus, our approach provides a useful framework to analyze the evolution and institutional changes of a monetized economy.

Proposition 5 is consistent with the literature stressing the importance of transaction costs in the evaluation of economic activities (e.g., Coase [1937], Williamson and Winter (eds.) [1993]). Unfortunately, much efficiency analysis has neglected the role of frictions in transfers. For example, standard general equilibrium analysis (e.g., Debreu [1959]) implicitly assumes zero transaction costs. In this context, money does not play any significant role: it is simply a commodity used as a numéraire good. As such, it is neutral in the sense that it does not play any role in resource allocation beyond choosing a reference unit of value (e.g., Shit [1997]). However, in the presence of transfer frictions, money is no longer neutral. First, money can help improve efficiency by reducing transaction costs. Second, money can contribute to efficiency gains even in the absence of exchange. As discussed in section 2, the choice of money was crucial for Coach Carter: he had to find a private good with contractible use rights that would be easy to monitor. His choice of pushups as debt payments was clever: no other good possessed such useful characteristics. In this context, the nonneutrality of money is related to its ability to help solve a coordination problem. In general, this ability depends on the identification of contractible rights and on the monitoring technology.

Equation (11a) also illustrates the role of transfer cost in efficient allocations. It shows that efficient transfers $t$ are the ones that maximize the aggregate value of transfers, net of transfer cost $C(p_{\varphi}^t, p_f^t, t, \varphi, E)$. This is a form of the Coase theorem applied to exchange activities. Equation (11a) shows how monetary institutions $E$ affect the profitability of exchange, $\pi_z(p_{\varphi}^t, p_f^t, E)$. Again, comparing two monetary institutions $E \in \Omega$ and $E' \in \Omega$, finding that $\pi_z(p_{\varphi}^t, p_f^t, E') \geq \pi_z(p_{\varphi}^t, p_f^t, E)$ means that institutions $E'$ are more effective (than $E$) in facilitating exchange. In addition, for a given $E' \in \Omega$, equation (11a) provides useful insights into the efficiency of exchange. Assuming that $C(p_{\varphi}^t, p_f^t, t, \varphi, E)$ is differentiable in $t$, the Kuhn–Tucker conditions for $t_j \geq 0$ in (11a) are $p_{\varphi}^t - p_f^t \leq 0$, $t_j \geq 0$, and $[p_{\varphi}^t - p_f^t - \partial C/\partial t_j] \cdot t_j = 0$. This implies that, when $t_j > 0$, efficient transfers must satisfy $p_{\varphi}^t - p_f^t = \partial C/\partial t_j$, i.e., that the difference in shadow prices between agent $i$ and agent $j$ must equal the marginal transfer cost $\partial C/\partial t_j$. This has two implications. First, it shows that the shadow prices $p_{\varphi}^t$ can be uniform across agents only if the marginal costs of transfers are zero. In other words, the law of one price $(p_{\varphi}^t = p_f^t$ for all $i, j$) can

---

18 Note that the same thing can be said of the reference bundle $g$ used in sections 4 and 5 to characterize efficiency. As noted above, the reference bundle $g$ is used simply as a convenient means of evaluating the efficiency of resource allocation. As such, the reference bundle $g$ should not be identified with money.
apply only in the absence of transaction costs. Second, frictions in transfers cause the shadow values $p_i$ to vary among agents. This variation can make it difficult to discover the shadow values of goods.

Finally, note that the maximization of values in (11a) applies at the aggregate level. First, equations (11) are conditional on $p$, the set of firms $J$, and on the technology $Z(p, E)$. Being conditional on $J$ implies that the analysis depends on the structure of production and the underlying property rights. Second, equations (11a)–(11b) show that in the presence of externalities in transfers across agents, the choice of $z$ and $t$ cannot be decentralized. For example, the presence of network externalities in trade can create significant challenges to decentralized exchange.

This highlights the limitations of explaining the historical emergence of money simply as a consequence of the voluntary individual choices taking place in market exchange.

Under the convexity of $Z(p, E)$, equations (9), (10), and (11) provide a characterization of a maximal allocation, conditional on the shadow prices $p$. This raises the question: How do we know the value of $p$? The Kuhn–Tucker conditions associated with the Lagrangian (8) give the complementary slackness condition

\begin{align}
\tag{12a}
& p_i^f \cdot \left[ y_i - z_i - \sum_{i \in N \cup J} t_{ij} \right] = 0, \quad p_i^f \geq 0, \quad \left[ y_i - z_i - \sum_{i \in N \cup J} t_{ij} \right] \geq 0, \quad i \in N \cup J, \\
\tag{12b}
& p_i^m \cdot \left[ \sum_{i \in N \cup J} t_{ij} - x_{ai} \right] = 0, \quad p_i^m \geq 0, \quad \left[ \sum_{i \in N \cup J} t_{ij} - x_{ai} \right] \geq 0, \quad i \in N \cup J, \\
\tag{12c}
& p_b \cdot [y_b - x_b] = 0.
\end{align}

This gives the standard results that prices are positive whenever increasing the availability of the corresponding commodities improves aggregate welfare. In the context of market economies without transfer costs, the shadow values $p$ can be interpreted as the competitive market-clearing prices. However, in the presence of transfer costs, externalities, and public goods, discovering the shadow values $p$ can be challenging. First, externalities and public goods imply that decisions cannot be easily decentralized. Second, in the presence of frictions in transfers, the shadow values typically vary among agents. Finally, as discussed next, information costs make it difficult to identify efficient decisions.

6.2 The Role of Information

Under a state-contingent approach to uncertainty, goods are interpreted to be state-contingent, i.e., conditional on particular states of nature. As noted above, this allows for a general representation of uncertainty, including asymmetric information (where different individuals face different partitions of the state space). And in a dynamic context, state contingency allows for information-gathering activities and learning. In general, the information-processing technology can involve externalities and
nonconvexity (RADNER [1968]). Note that under such situations, zero-maximality remains a valid representation of efficiency.\(^{19}\) In this context, efficient information $$\varphi$$ as well as efficient monetary institutions $$E \in \Omega$$ maximize the aggregate benefit in equation (4), with $$u$$ being chosen to satisfy $$V(u) = 0$$.

To illustrate the role of information, consider a feasible improvement in information technology from $$T(\varphi, E)$$ to $$T'(\varphi, E)$$ (where $$T(\varphi, E) = Y(\varphi, E) \times Z(\varphi, E)$$), with $$T(\varphi, E) \subset T'(\varphi, E), E \in \Omega$$. This change could be due to improvements in the ability of decision-makers to process information. Or, under asymmetric information, it could be due to a feasible reassignment in decision-making (who decides what) toward better-informed individuals − meaning that (at least some) state-contingent decisions are now associated with a finer partition of the state space. Importantly, the expansion in the feasible set $$[T'(\varphi, E) - T(\varphi, E)]$$ depends on the exchange and monetary institutions $$E \in \Omega$$. This suggests the potential for significant interactions between information and market institutions. To see this, after extending equation (4) to include maximization with respect to $$E \in \Omega$$, denote the surplus associated with $$T(\varphi, E)$$ by $$V(u)$$, and the surplus associated with $$T'(\varphi, E)$$ by $$V'(u)$$. With $$T(\varphi, E) \subset T'(\varphi, E)$$, it follows from (4) that $$V'(u) \geq V(u)$$. In other words, improvements in information technology create efficiency gains. When these efficiency gains are associated with a reassignment in decision-making, this means that modifying the distribution of rights and duties among individuals contributes to efficiency gains. This shows how the design of property rights and contracts is an integral part of economic efficiency.

Our analysis has two implications. First, as information technology improves, significant interactions are likely to arise between monetary institutions and property rights. For example, while a limited set of property rights may be adequate in a barter economy, the efficiency of a monetized economy exhibiting extensive exchange requires refined property rights. This suggests that the effects of a particular monetary policy will vary significantly with the underlying structure of property rights. Second, one expects that a reduction in imperfect information will help improve welfare. This suggests that transparency in monetary policy is, in general, efficiency-enhancing.\(^{20}\) This supports the recent trend toward greater transparency in monetary policy (e.g., BERNANKE [2003], GOODFRIEND [2007]).

As noted above, in the presence of externalities and public goods, there is a need to develop coordination schemes that can implement efficient allocations. We now explore the implications of uncertainty and information for the management of coordination schemes, including the choice of exchange and monetary institutions. In this context, the information requirements necessary to implement efficient allocations

\(^{19}\) This means that maximality as given in (4) applies under uncertainty, asymmetric information, nonconvexities, and externalities.

\(^{20}\) However, in a second-best world, the implications of information for a monetary economy can be unclear. For example, KAHN, MCANDREWS, AND ROBERDS [2005] have argued that, in an economy with imperfect contact enforcement, money can generate welfare gains when its “privacy” reduces the adverse effects of opportunistic behavior.
can be quite demanding. This is particularly true in complex environments where the number of states $S$ is large. In this context, it seems reasonable to consider situations of bounded rationality, reflecting the difficulties decision-makers have in obtaining and processing information. Note that our model captures this by allowing for costly active learning. As noted above, this corresponds to $T(\mathcal{G}, E) \subset T(\mathcal{G}', E)$ when the information structure $\mathcal{G}$ is at least as fine as $\mathcal{G}'$. When the economic environment becomes very complex, it seems reasonable to assume that $T(\mathcal{G}^+, E) = \emptyset$, where $\mathcal{G}^+ = \{P^+, P^+, \ldots\}$ reflects perfect information. This reflects bounded rationality, where making all decisions under perfect information $P^+$ is not possible (Simon [1955]). In this context, the maximal allocation in (4) provides a framework to analyze efficient allocations under bounded rationality. Then, efficient information is given by the information structure $\mathcal{G}$ that maximizes (4), evaluating the benefit of better information (represented by the constraint (2) becoming less binding) compared to its cost (represented by a reduction in the feasible set $T(\mathcal{G}, E)$ reflecting the resources used in the learning process).

The key argument is that a centralized coordinator (e.g., Coach Carter) may be able to process the information required to implement an efficient allocation only when the economy is simple. In a complex economic environment, the number of states of nature can become very large. Then, limitations on the ability to process information can impose severe restrictions on decision-making (Simon [1955]). In this context, a centralized scheme would likely fail to provide efficiency-improving contracts and an efficient solution to a coordination problem.

Could more decentralized solutions be used instead? A useful way of thinking about it is in the context of evolutionary strategies that concern the endogenous emergence of money and market places (e.g., as argued by Hayashi and Matsui [1996] and Matsui and Shimizu [2005] in the context of monetary exchange). Consider a coordination problem in which there exists an efficiency-improving allocation that can increase aggregate benefit (sections 4 and 5). A decentralized solution to the coordination problem could arise if (at least) one individual in the economy: (1) identifies a feasible allocation that can increase aggregate benefit; and (2) finds a way of distributing the increased benefit in such a way that would not induce opposition by other agents. In this case, this individual (or group of individuals) would be an institutional innovator who can contribute to solving the coordination problem in a more decentralized fashion. Note that identifying ex ante who is the innovator might prove difficult. If true, this has two implications. First, achieving efficiency improvements can be a slow process, suggesting that both past and current economies may spend long periods of time facing inefficient allocations (when no innovator arises). Second, the process of institutional innovation may be somewhat unpredictable. Both implications apply to monetary innovations through the ages. Interestingly, this process of institutional innovation would eventually contribute to increasing aggregate benefit. Over the long term, this can converge to a maximal allocation and to a zero-maximal and efficient allocation (Courtault and Tallon [2000]). This means that decentralized innovations could eventually lead to efficient exchange and monetary institutions. Note that
MATSUI AND SHIMIZU [2005] obtained a similar result: in the context of monetary exchange, they show how the evolutionarily stable outcome of market institutions would lead to an efficient equilibrium. Our analysis extends their result by making it applicable to situations in which money plays a role beyond being just a medium of exchange. Indeed, when money plays a role in the design and enforcement of property rights associated with the management of coordination failures, our analysis shows that evolutionary strategies associated with decentralized schemes can help support a longer-term move toward efficiency. As documented in our review of the history of money, this process appears to be very slow. The main reason appears to be information difficulties associated with bounded rationality. This points to the limitations of analyzing money solely in a context of exchange and decentralized search. It also stresses the role of bounded rationality in the analysis of money.

7 Conclusion

We have explored the linkages between money and economic efficiency. The analysis is presented under general conditions to provide new insights into the multiple functions of money, with a special focus on its nonexchange role. We believe that the standard account of money simply in terms of its role in exchange obscures its historical origins, and also hides important insights into the determinants of economic efficiency.

Our approach is presented under general conditions allowing for frictions in transfers. This has several significant advantages. First, our analysis is consistent with the classical arguments that money plays a role in reducing transaction costs. Second, by considering both the benefit and cost of transfers, our approach treats the transfer decisions endogenously, thus providing useful insights into the economics of exchange. Third, our efficiency analysis applies with or without exchange. This allows money to play a nonexchange role in efficient allocation. We develop this argument with a special focus on the role of money in solving coordination problems associated with externalities or public goods. In this context, we argue that money has a nonneutral role in the functioning of an economy. This allows us to challenge the standard result of money neutrality obtained when money is seen only as medium of exchange. The insights obtained from our analysis apply under bounded rationality. This seems particularly relevant in situations exhibiting significant changes in property rights in the process of policy reform or structural transformations.

Money is at the heart of Coach Carter’s coordination problem (section 3). This is a situation in which the value of exchange is zero (since there was no exchange), while the improvement in team performance is large. In this case, money supported a debt contract designed and enforced by Coach Carter that induced better team performance. Human history is a full of struggles trying to blend the interest of individuals with the interests of their community. Whenever these two interests do not fully coincide, a coordination problem arises and money is needed in the design
and enforcement of property rights and contracts. Under such scenarios, money has a nonneutral role and contributes to creating economic benefits associated with these rights and contracts that help solve coordination problems. We believe that such benefits are large for most economies, irrespective of the amount of exchange. The evidence from limited-exchange economies comes from anthropological studies of primitive societies (e.g., EINZIG [1948], DALTON [1965], HODGES [1988]). And there is empirical evidence in modern economies showing that the benefits from competitive exchange are only a small part of total economic progress (LEIBENSTEIN [1966], KELLER [2000], PAVCNIK [2002], FERREIRA AND ROSSI [2003]).

Our exploration of money indicates that, through an evolutionary process, decentralized institutional innovations can eventually lead to more efficient allocations. This generalizes the analysis presented by MATSUI AND SHIMIZU [2005]. Whereas they focused on monetary exchange, we have investigated economic efficiency in situations where money plays a role beyond being just a medium of exchange. Indeed, when money plays a role in the design and enforcement of property rights associated with the management of coordination failures, we argue that evolutionary strategies associated with decentralized schemes can help support a longer-term move toward efficiency. Yet, when applied to the history of money, this process is very slow, presumably due to bounded rationality. Our analysis suggests directions for further explorations of how bounded rationality affects the economics of money.

Appendix

PROOF OF PROPOSITION 1 Consider a feasible allocation \((x, y, z, t, \wp)\). Assume that it is not a maximal allocation. Then, there exists another feasible allocation \((x', y', z', t', \wp')\) such that \(\sum_{i=1}^{n} B_i(x'_a, x'_b, u_i) > \sum_{i=1}^{n} B_i(x_a, x_b, u_i)\). Under Assumption 1, the increased aggregate benefit can be feasibly redistributed among the \(n\) individuals. When the individuals benefiting from the redistribution satisfy nonsatiation in \(g\), this would make these individuals better off without making any one else worse off, thus violating the Pareto efficiency criterion. It follows that a Pareto-efficient allocation must be maximal.

Next assume that \((x^*, y^*, z^*, \wp^*)\) is a maximal allocation. If it satisfies \(V(u) < 0\), then reaching the utilities \(u = \{u_i : i \in N\}\) is not feasible. Thus, feasibility requires \(V(u) \geq 0\). Assume that \(V(u) > 0\). Under Assumption 1, the positive aggregate surplus \(V(U)\) can be feasibly redistributed among the \(n\) individuals. When the individuals benefiting from the redistribution satisfy nonsatiation in \(g\), this would make these individuals better off without making any one else worse off. Again, this would violate the Pareto efficiency criterion. It follows that Pareto efficiency implies a zero maximal allocation, with \(V(u) = 0\).

21 On these issues, see GROSSMAN AND HART [1986] and HART AND MOORE [1990].
PROOF OF PROPOSITION 2 Consider a feasible allocation \((x, y, z, t, \wp)\). Assume that it is not Pareto-efficient. Then, there exists a feasible allocation \((x', y, z', t', \wp')\) such that \(U_i(x_{ai}, x'Bj) \geq U_i(x_{ai}, xBj)\) for all \(i \in N\) and \(U_j(x'_{bj}, x'Bj) > U_j(x_{bj}, xBj)\) for some \(j \in N\). Under Assumption 2, it follows from (3) that \(B_i(x_{ai}, x'Bj, U_i(x_{ai}, xBj)) \geq B_i(x_{ai}, xBj, U_i(x_{ai}, xBj)) \geq 0\) for all \(i \in N\) and \(B_j(x'_{bj}, x'Bj, U_j(x_{bj}, xBj)) > B_j(x_{bj}, xBj, U_j(x_{bj}, xBj)) \geq 0\). This implies that \((x, y, z, t, \wp)\) cannot be zero-maximal.

References

BERNANKE, B. S. [2003], “A Perspective on Inflation Targeting,” Annual Policy Conference of the National Association of Business Economists, Washington, D.C.
EINZIG, P. [1948], *Primitive Money*, Eyre and Spottiswoode: London.


