Endogenous money: the evolutionary versus revolutionary views*

Louis-Philippe Rochon
Department of Economics, Laurentian University, Ontario, Canada

Sergio Rossi
Department of Economics, University of Fribourg, Switzerland

The purpose of this paper is to shed light on the endogenous nature of money. Contrary to the established post-Keynesian, or evolutionary, view, this paper argues that money has always been endogenous, irrespective of the historical period. Instead of the evolutionary theory of money and banking that can be traced back to Chick (1986), this paper puts forward a revolutionary definition of endogenous money consistent with many aspects of post-Keynesian economics as well as with the monetary circuit approach. This alternative starting point leads to the conclusion that money is indeed ‘always and everywhere’ an endogenous phenomenon. This analysis rests on the nature of debt, the essential role of a settlement institution, and the link between production, money and income.

Keywords: banks, endogenous money, monetary analysis

JEL codes: E42, E51, E52

1 INTRODUCTION

Economists interested in contemporary monetary theory mostly find themselves being confronted by two diametrically-opposed views, which, following Schumpeter (1954 [1994], p. 277), we may label respectively as real and monetary analysis (see Rogers 1989, ch. 1). According to the first view, largely associated with neoclassical or mainstream economists, money is merely a veil behind which the real economy operates, unhampered by monetary factors: ‘money enters the picture only in the modest role of a technical device’ (Schumpeter 1954 [1994], p. 278). For these economists, money plays no role except in determining nominal magnitudes. This means that the real and the monetary sides of the economy are independent of each other, notably in

* The first draft of this paper was presented at the European Association for Evolutionary Political Economy 2004 Conference on ‘Economics, History and Development’ held at the University of Crete, Greece, on October 28–31, 2004. Sergio Rossi carried out further research work while he was the Commerzbank guest professor of monetary economics at Chemnitz University of Technology, Germany (April–May 2005), to which he is grateful for having provided such an ideal working environment to carry out research work. The authors would like to thank Philip Arestis, Giuseppe Fontana, Claude Gnos, Peter Howells, Marc Lavoie and Virginie Monvoisin for their helpful comments. They are also indebted to Nunzio F. Canova and Dante A. Caprara for bibliographic assistance. The usual disclaimer applies.
the long run, although in the short run money may play a role, but partisans of real analysis usually consider this role as the result of some market frictions or imperfections, which money would help ‘lubricate’.

By contrast, according to proponents of monetary analysis – which notably include post-Keynesians, but also supporters of the monetary circuit approach as well as a number of Sraffians, Institutionalists and Marxian scholars – the existence of money is paramount in economic analysis. As a matter of fact, it is impossible to study economics, in particular macroeconomics, without first understanding the conception of money, its logical origin and creation, and how money is linked to production and income. For this second group of economists, the supply of money is endogenous, because it is determined by the agents’ demand to pay for the costs of production. As such, money is the means of payment that is required to settle any kind of debt obligation finally. As Lavoie (1996, p. 533, emphasis in original) explains, ‘this endogeneity is not a matter of institutions but rather one of logical necessity’. As such, money must be integrated at the very beginning of the analysis, or, as Schumpeter (1954 [1994], p. 278) tells us, ‘on the very ground floor of our analytic structure’. It is indeed necessary from a logical as well as factual point of view to measure and settle debt obligations arising within society. This requires disposing of a means of final payment, independently of the existence of uncertainty as regards the economic system within which agents carry out their transactions in any kind of market (factor, goods or capital).

Now, it is true that a number of mainstream economists have advocated some form of endogenous money. Wicksell is a prime example of this, and is currently undergoing a revival of sorts with the recent rise of the so-called New Consensus (New Keynesianism mark II) – see, for instance, Kriesler and Lavoie (2007), Rochon (2007) or Setterfield (2007). Yet Wicksell maintained the existence of a natural rate of interest, as do proponents of the New Consensus, which clashes with the post-Keynesian rejection of this natural rate concept. As another example, Milton Friedman also believed that money is endogenous in an open economy with fixed exchange rates. Yet, as Lavoie (1996) and Rochon (1999) argue, these views on endogeneity stand in contrast to the post-Keynesian view. For mainstream economists, while the money stock is of an endogenously determined magnitude, there is still some ultimate supply-determined stock of money, which, at the end of the day, can be rationed by the central bank. In fact, recent discussion about the minimal success of so-called ‘quantitative easing’ (QE) policies shows precisely our point. Mainstream economists are somewhat bewildered at the failure of QE policies to promote output growth. Indeed, QE is ultimately based on the idea that money is exogenous and on the money multiplier framework: if the central bank somehow increases the liquidity of banks, the latter would in turn increase their lending and get the economy growing again. Yet, from a post-Keynesian perspective, the ‘failure’ of QE should have been anticipated all along: banks do not lend simply because they have increased liquidity.

Among monetary analysts, we would argue, only post-Keynesian authors have made, so far, endogenous money the cornerstone of their monetary theory of production. Indeed, money’s endogeneity is embedded in their theory of output and growth, and regardless of the horizontalist–structuralist debates of some decades ago, there is today much wider consensus, largely because of the circumstances of the day (see for instance Lavoie 1996 for a discussion). In particular, there is consensus on the following four important points. First, post-Keynesians agree that the ‘supply of money’ is demand-determined in order to finance the needs of the economic system to reproduce itself and grow, through the use of bank credit: money is issued through bank credit. Second, there is also agreement that banks are special because they can issue any money units, and that their role is of
primary importance in economic activity. Banks are not constrained by prior central bank reserves, and lend to all customers they deem creditworthy, although what they consider as creditworthy may change over the cycle (see Rochon 2006). Third, there is also some agreement on the role of central banks in preventing financial crises by supplying banks with the liquidity the latter need. Indeed, the crisis in Europe has exemplified the important role a central bank can play in mitigating the impact of a crisis. The refusal of the European Central Bank, for instance, to intervene more aggressively has had some devastating effects. Central banks must act as a lender of last resort. Finally, there is now, we would argue, also a large consensus on the exogenous nature of the central bank controlled rate of interest (see Rochon and Rossi 2006). Central banks can determine any rate of interest they wish, as is clearly the case with near-zero interest rates in several countries at the time of writing.

This consensus aside, the aim of this paper is to explore a remaining point of debate among post-Keynesians: it concerns the historical account of money’s endogeneity. In other words, the question we wish to explore and answer here, from an analytical as opposed to a historical point of view, is whether money has always been endogenous, or whether its endogeneity is the result of recent events and, in particular, of the evolution of the monetary and financial system. Needless to say, post-Keynesians have always been interested in this topic, and the issue that we discuss in this paper was always implicitly in the background of post-Keynesian analyses. For this reason, within post-Keynesian literature we can identify two overall approaches to this question, which we label here the ‘evolutionary’ and ‘revolutionary’ views respectively.

For evolutionary post-Keynesians, money’s endogeneity is a result of a long process and evolving institutions. In this sense, money only became endogenous quite recently, as the result, in particular, of some financial innovations and/or an accommodating central bank (Chick 1986; Pollin 1996). Before then, the orthodox story of money would apply. Hence, money became endogenous through time and in particular through the evolution of the banking system.

For revolutionary post-Keynesians, however, money has always been endogenous, irrespective of the historical period or of specific institutional arrangements. This position is advocated, among others, by Lavoie (1996, p. 533), who argues that ‘accommodation or the lack of it, liability management or the lack of it, and financial innovations or the lack of it are second-order phenomena compared to the crucial story that goes from debt creation to the supply of means of payment’. In other words, the money creation process is an endogenous phenomenon irrespective of central bank behaviour, the stage of development of the banking sector, financial innovations or other recent institutional changes.1

Now, while this paper sets out to explore these two approaches to money’s endogeneity, it is important to emphasize that there are, as stated above, some strong, common elements to both approaches, and that the consensus amongst post-Keynesians is that money is today an endogenous phenomenon. Nevertheless, the discussion presented in this paper may be important for both theoretical and policy reasons, a point that we address later. Overall, however, our position is that money has always been endogenous and, in this sense, we sympathize with the revolutionary approach to money’s

1. Rochon has recently argued (2012), however, that while this is correct, a full discussion of endogenous money does involve the central bank, otherwise it can lead to a New Consensus-type of endogeneity, where banks create money, but are somehow constrained by central bank activities. In this revolutionary approach, the central bank fully accommodates banks’ needs always and everywhere.
endogeneity. We believe that showing that the money creation process has always been endogenous is important in particular for post-Keynesians, since it will help to put the role of central banks and other monetary institutions, as well as financial innovations, in their proper theoretical perspective.

The structure of this paper is as follows. The next section briefly overviews the essential elements of endogenous money theory as it now stands. The third section discusses the so-called evolutionary view of endogenous money, as argued by Chick’s (1986) theory of the evolution of the banking system. In its stead, we propose in the fourth section what we call the revolutionary2 (see Rochon 1999) definition of endogenous money consistent with many aspects of post-Keynesian economics as well as with the monetary circuit approach, where the money creation process is always and everywhere an endogenous phenomenon, irrespective of the historical period, and therefore even under the gold-standard system. This alternative, analytical view rests on the nature of debt, the role of settlement institutions, and the functional link between production, money and income. Indeed, the historical evolution of the banking system has simply been revealing and making plain what has always been the case in actual fact: to wit, the endogenous nature of money emissions. The fifth section concludes.

Before addressing our topic, let us note that, in order to do justice to the ideas presented in this paper, we would need much more space to develop them more comprehensively on historical grounds, and to extensively articulate the arguments behind them. In spite of the space constraints imposed herein, however, in this paper we develop these ideas to a sufficient extent in order to start off a promising discussion among endogenous-money proponents so as to strengthen monetary analysis on theoretical grounds. In fact, the ideas that we challenge here are crucial for endogenous money theory and policymaking, and this paper provides the theoretical foundations for a much-needed debate that should clarify, and hopefully enlighten, the heterodox meaning of money’s endogeneity. Let us hope that other post-Keynesians will take up this challenge and provide further evidence on the historical roots of money’s endogeneity in the not-too-distant future. As a matter of fact, our article is not a historical paper, in the sense that it does not deal with the history of economic facts. Rather, it is a theoretical paper dealing with monetary analysis. In so doing, it provides nevertheless a few historical examples to illustrate and to buttress our arguments.

2 ENDOGENOUS MONEY: A QUICK OVERVIEW

Endogenous money is at the centre of post-Keynesian macroeconomics.3 It is a tradition that began in the post-Keynes era with Joan Robinson (1956), and continued with Kaldor

2. The ‘revolutionary’ essence of the definition of endogenous money proposed in this paper is tied to the spirit of Keynes’s ‘revolution’, namely his attempt to break away from orthodox thought completely. It is a paradigm change with respect to neoclassicism, and in particular neoclassical monetary economics. As a matter of fact, most horizontalists, especially those of the monetary circuit approach, do not agree with Chick’s stages approach to money’s endogeneity. This group includes, among possible others, Graziani, Lavoie, Rochon and Seccareccia. Note that, among structuralists, Wray would probably also side with the revolutionists.

3. See, for instance, Moore (1988), Wray (1990), Arestis (1992), Lavoie (1992), and Rochon (1999), among a great many others. A number of scholars have set out to verify empirically endogenous money theory. Among them, Shanmugam et al. (2003) look at Malaysia; Palacio-Vera (2001) studies the case of Spain; and Howells and Hussein (1998) consider the G7 countries. See also Arestis and Sawyer (2003) in this respect.
(1970), Eichner (1987), and Moore (1979a; 1979b; 1983; 1988). Today, it is elaborated and deeply embedded in post-Keynesian economics, and stands as a rallying point of departure from many mainstream approaches. As post-Keynesian scholars explain, money appears in the economy along with production when banks agree to honour debt contracts with firms (Davidson 1972). As the economy grows, banks supply more credit to meet the growing needs of the system, either to pay out wages or to purchase any factors of production. As soon as a firm receives credit, it can pay out wages: at this very instant, money is created by the bank carrying out the payment. The creation of money is thus parallel to, but must not be confused with, the creation of income. We may thus claim, as argued by Robinson (1956) and post-Keynesians in general, that the supply of money expands and contracts with the needs of production, in response to expectations of aggregate demand, through the banking system (see Arestis and Eichner, 1988). This is the essence of the theory of endogenous money as it now stands today.

Even in a stationary system there is always a need to supply bank credit, as firms must reimburse the credit they need to finance production. At the beginning of each production period, in fact, a new credit must be secured by firms, and banks must agree to supply this new credit or to roll over existing debt (Rochon 1999; Parguez and Seccareccia 2000). Of course, banks may always refuse to grant credit, in light of bad profit expectations or bad expectations on the future course of the business cycle. Hence, banks are never passive in their lending decisions, and there may always be a ‘fringe of unsatisfied customers’ (as Keynes put it). When banks refuse to grant credit, the economy can face a severe liquidity shortage, which might lead to a downturn in the business cycle and

4. Recently, a number of mainstream economists of the New-Keynesian perspective have acknowledged the endogenous nature of money, noticing that the rate of interest is exogenous and that the banking system creates money at the request of borrowers. Under the auspices of the so-called New Consensus model, which includes a Taylor-rule-based policy reaction function for the central bank, the similarities to the post-Keynesian approach are nevertheless limited. Indeed, while adherents to the New Consensus recognize the endogenous nature of money, they still believe in the long-run neutrality of money. Moreover, rather than seeing endogenous money as a natural occurrence in capitalist economies of production, they advocate the endogeneity of money on the grounds that the latter is the result of preferred policy, owing to the instability of the demand-for-money function. In other words, while New Consensus proponents have recognized the endogenous nature of money, they do not have a theory of endogenous money. In fact, as Arestis and Sawyer (2004, p. 442, fn. 2) clearly put it, the New Consensus approach considers ‘money as a residual with no further role for it [while the post-]Keynesian notion of endogenous money entails a fully articulated theory with clear policy implications where money and credit have important roles to play in their interaction with real variables’.

5. We leave undiscussed the issue of the financing of investment, as this is an issue of some disagreement among endogenous money proponents (see Seccareccia 2003; Messori and Zazzaro 2005; and Rochon 2005).

6. Owing to space constraints, we do not address the roles of households and the State in this story. See, for instance, Wray (1998) and Rochon (1999) for elaboration on these roles.

7. Readers will notice that there is no reference here to a central bank. Indeed, as argued by some horizontalists, there is no need for a central bank to explain the endogeneity of money. Money is a means of final payment, irrespective of the existence of a central bank. For a complete picture of the process of money creation and destruction, and the validation of a bank’s acknowledgement of debt, a central bank would be needed; but none is required to explain the endogenous nature of money. This is what Lavoie (as quoted above) meant by endogeneity being a matter of ‘logical necessity’.
to higher unemployment. As long as firms are deemed creditworthy, however, they will receive the necessary credit. Similarly, as banks are deemed creditworthy, they should face virtually no limits to the amount of credit they can provide as long as they ‘move in step’ with other banks, as Keynes (1930 [1971], p. 23) pointed out.8

If this view is correct, then what about ancient monies? Can the same logic be applied to them? For instance, was money already endogenous in 750 BC when the people in Lydia (Asia Minor) were inventing the first coins,9 or in 400 BC when the Mesopotamians and Egyptians were instituting earlier forms of banking? In other words, was money already endogenous in ancient times, or did it become endogenous over time, as Chick (1986) and many other post-Keynesians argue, in particular with respect to the creation of central banks?10

3 THE EVOLUTIONARY VIEW: A CRITICAL APPRAISAL

In a widely-quoted paper, Chick (1986, p. 113) argues that, in the early stages of banking, the money creation process was entirely exogenous. Banks were neoclassical, in the sense that they were “‘conduits’ between saving and the employment of those savings for investment […] saving determines the volume of investment’. At that time, according to Chick (1986), the causality ran from bank deposits to reserves, and finally to bank loans. The supply of credit was therefore pre-determined, and the principle of scarcity applied. Banks were pure financial intermediaries in this view. This was the first stage of financial development.11

Stage two, in Chick’s (1986, p. 114) view, is when the “‘bank deposit multiplier’ is the relevant theory: the banking system can now lend to a multiple of reserves, subject to a conventional or imposed reserve requirement’. It is only when the central bank has ‘fully accepted responsibility for the stability of the financial system’ (Chick 1986, p. 115), in the fourth stage of banking development, that banks were able to expand lending beyond their reserve capacity. It would then be the accommodative role of the central bank – that is, the removal of reserve constraints – that rendered money endogenous, in the loans-to-deposits causality sense. Endogenous money is thus viewed as the result of institutional changes, defined as the ability of the banking system to expand the supply of credit with no prior expansion of bank reserves. Stage five, the introduction

8. Lavoie (1992, p. 201) argued that as long as banks ‘move forward in step’ with the lending activities of other banks, they can lend as much as they want. Reference is made here to Keynes’s Treatise on Money, in which he argued this point precisely. See also Graziani (2003, p. 63) for a restatement of the same argument in justifying a central bank’s role. Note incidentally that this argument was already made by Le Bourva (1962), whose English translation was published 30 years later (Le Bourva 1992).

9. ‘According to Xenophanes, our earliest authority, coinage was invented by the Lydians. Herodotus, our next authority, seems to concur. The earliest coins have been found in western Asia Minor, particularly in Ionia and Lydia’ (Cook 1958, p. 261).

10. We do not discuss here the chartalist approach to money (see Wray 1998), which explains the creation of money and its purchasing power by referring to the existence of taxes that the population has to pay to the local State. See, however, Rossi (1999), Gnos and Rochon (2002) and Rochon and Vernengo (2003) for a critical discussion of the chartalist view.

11. The success of this argument within post-Keynesian circles remains a mystery. Indeed, Chick (1986 [1992], p. 193) explains that the evolutionary scheme she used in her 1986 article was ‘invented during a graduate student’s supervision’. This might suggest that her stages hypothesis was put to the fore ex nihilo, without any preliminary historical investigation.
of liability management, is considered a less fundamental version of money’s endogeneity, when banks actively seek to pursue financial innovations in order for them to expand the supply of credit. This last stage arises only because the central bank may not fully accommodate the demand for bank reserves. In both these stages, however, the result is essentially the same: banks no longer depend on deposits or reserves in order to lend.

The main conclusion of this analysis is that money is endogenous or exogenous according to the historical period considered. This argument has been advocated by a number of post-Keynesians. Minsky (1991, p. 208), for instance, claimed that ‘there are periods in history and economic conditions where the money supply was mainly endogenous and other periods and conditions where the money supply was largely exogenous’ (see also Davidson 1972; Davidson and Weintraub 1973; Dow 1988; Guttmann 1990; Niggle 1991). Moore (1996, p. 89) claims in this connection that the exogenous/endogenous debate is somewhat misleading, since ‘both views are correct, but […] each is applicable to a different historical period’. Even Eichner (1987, p. 849) seemed to hold such views: ‘The reason the earlier systems of commodity and fiat money alone were replaced by a credit-based system was the need for a means of payment that would vary with the level of economic activity and thus with the need for additional funds’.

Many post-Keynesians, as stated above, adhere to this evolutionary view. The question to raise in this respect is ‘why?’, especially in light of anthropological evidence suggesting that the money creation process has always been an endogenous phenomenon in the history of mankind. One possible explanation can rest in what Monvoisin (2006) has called the two poles of money endogeneity. As stated earlier, while there is no need for a central bank to explain the endogenous nature of money, a central bank is required nonetheless to get a full picture of the circuit of money and cancellation of bank debt. Money’s endogeneity can therefore be explained by referring to two crucial relations: the relation between banks and firms, and the relation between banks and the central bank. Indeed, most post-Keynesian authors have traditionally emphasized the second relationship, placing much less emphasis on the first relationship (which is traditionally emphasized by proponents of the monetary circuit approach, for example). In fact, many post-Keynesians identify the endogenous nature of money with central banking, as depicted for instance by Chick (1986). Rousseas (1989, p. 478) summarizes this position clearly: ‘The degree to which the supply of money is positively sloped depends on the discretionary policies of the Federal Reserve’. This is, in our view, a misconception of the endogenous nature of money, as the next section will show on analytical grounds.

4 THE REVOLUTIONARY VIEW: SOME FIRST PRINCIPLES

The purpose of this section is to challenge on analytical grounds the evolutionary view and the premise that money was at one point in history an exogenous variable controlled by monetary authorities, and that, through time, owing to the development of banking and financial innovations, somehow money became out of reach for monetary authorities. To make our point, we need to show that ‘money’ predates the usage of coins. In this sense, there is no relationship between money and the money supply conventionally understood. ‘Money’ goes along with the needs of production, as any payments of production costs need a means to settle debt obligations finally. Hence, in this sense, money has always been responding to the needs of markets for a means of final payment, and has therefore always been endogenous. If so, the endogeneity of money is not a recent
phenomenon. Indeed, ‘the orthodox sequence of “commodity (gold) money” to credit fiat money does not square with the historical record’ (Wray 2006, p. 6).

Interestingly enough, some post-Keynesians seem to have reached this conclusion already, but on historical rather than logical grounds. For instance, as pointed out by Arestis and Howells (2002, p. 4) referring to the ‘great inflation’ in the United Kingdom between 1520 and 1640, endogenous money might have already existed in the distant past: ‘[t]here appears to have been some scope for [money’s] endogeneity, even in the very earliest “stages of banking”’. As a matter of fact, as noticed by the authors (2002, p. 9), at the time of commodity money only a fraction of the precious metal available (in the sense of having been mined and melted) was used as a circulating medium. According to Innes (1913, p. 389), ‘so small was the quantity of coins, that they did not even suffice for the needs of the Royal household and estates which regularly used tokens of various kinds for the purpose of making small payments. So unimportant indeed was the coinage that sometimes Kings did not hesitate to call it all in for re-minting and re-issue and still commerce went on just the same’. Courbis et al. (1991, p. 321) add another relevant point in this (historical) regard, namely that trade between distant cities did not occur on a commodity-money basis, owing to distance, geography, and the intensity of exchanges.

Courbis et al. (1991, pp. 321–322) further point out that in maritime trade, which was the dominant form of trade between cities in Europe and elsewhere, gold bullion and coins were not used, as a general rule, to settle debts, because of weight and space problems in shipping cargos that were much more useful for transporting real goods. To this evidence, Cook (1958, p. 260) adds that ‘the Phoenicians and other people of the East who had commercial interests managed satisfactorily without coined money’. Further, ‘[i]n the Greek world of the seventh and sixth centuries [BC] it is hardly likely that merchants would have had more trust in coins, especially if (as must often have happened) the coins offered were those of another state and of another standard. So it seems reasonable to suppose that coinage cannot have been invented to ease the larger commercial transactions’ (ibid., p. 260).

These historical arguments converge in showing that the amount of circulating medium was insufficient to sustain the level of production and economic activity. This might imply that several transactions were carried out using, say, goldsmiths’ debt certificates – that is, fiat money. In addition, because of foreign trade, gold inflows did not arrive like helicopter money within the country: ‘If a country runs a balance of payment surplus, it must expect (in the sixteenth and seventeenth century) to receive the surplus in specie’ (Arestis and Howells 2002, p. 9). As a consequence, if the endogeneity of money means that money supply is demand-driven, ‘even commodity money can provide an example of monetary base endogeneity, increasing in response to trade demands’ (p. 9). Money would thus be endogenous independently of its material form – hence independently of the historical period considered.

Now, for some post-Keynesians credit and money can never be constrained (see Lavoie 1992) – except perhaps by the demand for them. This point was indeed well recognized by Joan Robinson (1956). As her views on endogenous money are becoming increasingly considered (see Graziani 1989; Lavoie 1999; Rochon 2001; Gnos and Rochon 2003), we now find evidence that she also was of the opinion that money was endogenous even under the gold-standard system. As a matter of fact, if there were some physical limits imposed by gold mining, banks would simply step in and create the necessary amount of money in order to meet the needs of the economy for a means of final payment. Hence, even under the gold standard, the money supply adapted
to its demand. As Robinson (1956, p. 32) put it, ‘[g]eology being limited, the [gold] stock above ground did not grow nearly as fast as the demand for liquid balances, and banks came into existence to supplement the supply’.

The above statement by Robinson leads us to two immediate conclusions. On the one hand, as stated above, the money creation process is endogenous irrespective of the historical period, hence even during the gold-standard era (see below). In other words, the endogeneity of money is not linked to any specific institution, and predates the creation of banks and their development. On the other hand, there would have been no ‘stages of development’ of the banking system essentially. Banks would have been created for the sole purpose of settling debts between agents. In this sense, banks were never financial intermediaries, directing saving to investment, but rather money and credit providers, as Keynes pointed out in early drafts of his Treatise on Money.

In fact, the first, decisive step in the institution of banking was the introduction of double-entry bookkeeping. Made possible by the discovery of negative numbers (attributed to the Indian mathematician Brahmagupta in the seventh century), double-entry bookkeeping was developed only much later – in the thirteenth century – by Italian traders, who took advantage both of Arabic numerals and of the Indian conception of zero. It is indeed to Indian mathematicians that we owe the knowledge that zero has a definite numerical value of its own, that it is an even number (the integer that precedes one), and that it separates positive from negative numbers. Italian merchants were the first to realize that positive and negative numbers could be used to represent commercial transactions as well as to allow for the activity of newly-born institutions that became the ancestors of today’s banks (Cencini 2005, p. 299).

As Courbis et al. (1991, p. 332) point out, in fact, the account balance written on paper or merely recorded in the books of the ‘banks’ continued to be managed by merchants and was the result of credit activities. Hence, loans created deposits, which is tantamount to saying that the supply of money expanded with the demand for it. Further, Courbis et al. (1991, pp. 323–325) provide strong arguments according to which paper was discovered very early in history but only became a money form much later on: paper had been for a long time the material used by human beings

12. Lavoie (2003) explains that under the gold standard the money supply was not endogenously driven by supply-side considerations, such as balance of payment surpluses, as in the orthodox approach. He points out that any increases in the money supply caused by a trade surplus, for instance, are absorbed via the so-called ‘compensation effect’ when debts are reimbursed, thereby neutralizing such increases. These debts exist because either the banking sector is in debt towards the central bank, such as in overdraft systems, or because bank borrowers reimburse their existing debt towards the banks, as in the traditional post-Keynesian endogenous money view. In an asset-based system, banks would use these added reserves to purchase bills (rather than reimburse debts with the central bank). Note that these excess reserves cannot be used by the banks to increase lending, since loans can only be made at the request of borrowers, and banks would have already extended their loans to creditworthy borrowers. Under the gold standard, thus, the money supply was still determined by the demand for loans.

13. By way of contrast, as another example of the evolutionary view, many post-Keynesians see money as being exogenous under the gold standard. For instance, according to Moore (2001, p. 20), ‘[w]hen money was a commodity, such as gold, with an inelastic supply, the total quantity of money in existence could realistically be viewed as exogenous’.

14. If money is endogenous irrespective of the existence of a ‘bank’ as such, it nevertheless requires the intervention of a third agent. In the contemporary framework, this role is assumed by banks, but one can conceive of it being assumed by other agents, as we discuss below.
in order to keep accounting books in what were called at that time ‘banche di scritta’ (that is, accounting banks) in fourteenth-century Venice.

In Venice everyone ‘of consequence in business had an account so that he could make and receive payments through the banks. They were called banche di scritta or del giro because their main function was to write transfers and thus to rotate (girare) credits from one account to another.’ It is clear that these deposits on current accounts were a form of bank deposit money. (Copeland 1981, p. 252; citation from Lane 1973, p. 147)

This book-entry payment system was developed during the fourteenth century, probably in Genoa and Florence (Copeland 1981, p. 254, fn. 6). As Rostovtzeff (1941) claims, however, this system might have been used already in ancient Egypt about 300 or 200 BC: a second-century papyrus record (Teb 891.36) contains fragments of the daily payments recorded by a small branch bank in the Heracleopolitan province. ‘In many instances […] the payments were effected by transfer from one account to another without money passing’ (Rostovtzeff 1941, pp. 404–406). The same system was also in use in Greece at least by the fourth century BC (Copeland 1981, p. 249): when a debtor settled his debt obligation, his deposit account in the bank was debited and the creditor’s account was credited. According to Courbis et al., however, the double-entry bookkeeping system of banking reached its historical apex at the beginning of the seventeenth century with the Bank of Amsterdam. As they observe, ‘at that period, money is not made of paper; it is rather already book-entry money’ (Courbis et al. 1991, p. 325, our translation). If this view is correct, then this interpretation may strengthen the post-Keynesian theory of endogenous money by providing a firmer, and more heterodox, basis on which to build a general theory of a monetary economy of production. Let us elaborate on this point.

4.1 Money and debt

As the theory of the monetary circuit shows, although this is also a conclusion reached by post-Keynesians, money is a necessary result of a debt relationship between a borrower and a lender (see Graziani 2003, but also Hicks 1967, p. 11). In this sense, money is a social relation (see Ingham 1996). To place a value on the debt, a value system needs to be developed by society. In this respect, money is a social, numerical counter supplied by the banking system just as it is required, as was stated by Hicks (quoted by Laidler and Parkin 1975, p. 742). This then means that any transaction, and a fortiori any payment, necessitates a bookkeeping system to record debt obligations and their final settlement. In bookkeeping terms, the creation of money is the means by which the banking system provides the economy with a number of money units that are debited and credited to the payer, respectively to the payee, who use them to exchange objects between them.15 This number of money units is necessarily created by banks, or by their predecessors (such as goldsmiths). This creation may of course take different physical forms depending on the technological and institutional framework, but is always an endogenous phenomenon, as it stems from the agents’ demand for a means of final payment – be it in the form of gold coins, paper money, or purely book-entry money. As Ingham (1996, p. 510) argues, ‘all forms of money are social relations’, ‘including its archaic “commodity” forms’ (p. 525).16

15. Wray (1990) provides an insightful analysis of this approach. Unfortunately, this investigation has not been widely accepted or paid due attention by post-Keynesian scholars.
16. As Ingham (1996, p. 526) emphasizes, ‘it is not being suggested that distinctions between different forms of money should not be made; for example, commodity money, fiat money,
To be sure, it is beyond dispute that important changes have occurred in the material form taken on by money or in the way debt obligations are physically settled. However, unless we make the error of identifying money with what has been historically used to represent it, the changes in the material form of money cannot be taken as a proof of the evolution in its nature. Even if we accept the view that our economic systems are part of a ‘process of interdependent evolutionary change, a process which itself evolves’ (Chick and Dow 2001, p. 712), this does not imply that in such an open, evolving, and complex system as the actual economy, money itself is evolving in nature (that is to say, conceptually), as Chick and Dow (2001, p. 710) maintain. In this context, the evolutionary argument made by Chick and Dow (2001) would not stand for a change in the meaning of concepts, but for an increasing degree of conformity between the working of the economic system and the endogenous nature of money. More precisely, a positive evolution can occur only if the practical structure of the (monetary) economic system is made to comply with the set of macroeconomic laws implicit in the bookkeeping nature of money and its relation with production and exchange (Cencini 2005, pp. 280–281).

This conclusion can further be substantiated on historical grounds. Recall that in the Middle Ages, particularly at the beginning of the fourteenth century, great periodical fairs were being held all over Europe, the most famous of them in Champagne, France, to which came traders and bankers from everywhere. In order to settle transactions, ‘[e]xchange booths were established and debts and credits were cleared to enormous amounts without the use of a single coin’ (Innes 1913, p. 396) – so much so that ‘[a]t some fairs no other business was done except the settlement of debts and credits’ (p. 397). Such an institution, as a matter of fact, enabled merchants and traders to rely on a unit of account whose definition is stable, because of its being wholly independent of coin debasement – a practice that was not infrequent at that time. Clearly, these primitive forms of clearing houses defined a monetary space (see Ingham 2002), whose borders were set by the limits within which debts and credits were an endogenous phenomenon with respect to a given medieval fair.17

In fact, according to many historians, anthropologists and sociologists (see Polanyi 1944; Grierson 1977; and Ingham 2000, among several others), money originated in primitive societies as a unit of account and a means of payment. As a matter of fact, even in ancient, stateless societies, human relations implied a unit of account to measure and regulate the reciprocity of debt obligations as well as the redistribution of commodities. These human or social relations also implied a means of payment in order for individuals to settle their social debts, such as those arising from status, kinship, convention or religion (see for instance Malinowski 1921; Einzig, 1949 [1966]; and Polanyi 1977). In primitive societies, therefore, money did not appear as a market-induced result of a discovery process that cost-minimizing agents went through in order for them to avoid the drawbacks of the double coincidence of wants. Money was (and indeed is) based ‘on the antiquity of the law of debt’ (Innes, 1913, p. 391). In fact, what Innes (1913, p. 393) called ‘the primitive law of commerce’ is the basic principle of double-entry bookkeeping, which is to record debts and credits for further reference and

promissory notes, checks, credit cards, local exchange trading scheme (L.E.T.S.) tokens etc. Each has its own particular conditions of existence; but all such conditions are essentially social’.

17. ‘All money is created and maintained by the social relation of credit–debt […] These relations create the monetary space – that is, a social sphere in which impersonal exchange takes place’ (Ingham 2002, pp. 127–128).
settlement. From a technological point of view, this is tantamount to saying that ‘money is equivalent to a primitive form of memory’ (Kocherlakota 1998, p. 232). In other words, exchange as well as production requires a numerical system by which the value of these actions or transactions can be accounted. As Ingham (1996, p. 519) cogently puts it, ‘money as a measure of value is a “collective representation” for which the analogue is the structure of society’.18

Now, debt–credit relationships, and records, have neither logically nor historically to do with a particular material support, say gold. In other words, money’s purchasing power has no link with the stuff that carries out money’s function in both ancient and modern societies. This is a ‘first principle’ valid in any historical period, even the gold-standard era. To be sure, gold coins were just one possible form of money. They were largely used as a means of payment owing to the fact that the State had often a monopoly on the mints, and that with its imprimatur the State provided a universal guarantee against counterfeiting. However, the purchasing power of a coin under the gold standard was not driven by its metal content, weight, or backing. As Innes (1913, p. 382) pointed out 100 years ago, under the gold standard ‘the monetary standard was a thing entirely apart from the weight of the coins or the material of which they were composed. These varied constantly, while the money unit remained the same for centuries’. In fact, the purchasing power of a gold coin, as with the purchasing power of any other form of money, stems from the association of money with production activities, an association in which the monetary theory of production is firmly grounded (see Graziani 2003, ch. 3; Rochon and Vernengo 2003; Rossi 2003).

4.2 Money and production

In a monetary economy of production, ‘money is introduced into the economy through the productive activities of the firms, as these activities generate income. There can be no money without production’ (Lavoie 1984, p. 774). Let us focus therefore on the factor market and particularly on the payment of factor costs for produced output. This is indeed one of the fundamental building blocks of any monetary theory of production.

Consider a simple, stylized economy, in which a single bank operates. This, incidentally, may have some historical relevance with respect to ancient societies and moreover shows that the central bank (or its accommodative role) is not necessary in order for the money creation process to be endogenous. Let us refer to a simple numerical example, in which one primitive bank (say, a goldsmith) intervenes to ‘monetize’ the economy’s output. In any transaction where this bank intervenes, in fact, the latter issues the monetary form of the payment between two agents, namely, a payer, A, and a payee, B, say in the form of (a goldsmith’s) debt certificates. If so, then the goldsmith’s ledger records the transaction as depicted in Table 1.

Any transaction involves three parties, namely a payer, a payee, and a record keeper, that is, a ‘banker’ (as argued also by Hicks 1967, p. 11; and Parguez and Seccareccia 2000, p. 101). To be sure, if the payment were a ‘dyadic’ exchange – that is, a bipolar operation between the contracting parties – in which the payer handed over to the payee an acknowledgement of debt that the payer himself fabricates (say, a wooden tally), then the payment would not be final but merely promised. In fact, as argued by Graziani (2003, p. 60), ‘if a simple promise of payment could perform the role of final payment, buyers would be endowed with a seigniorage privilege, namely

18. See also Courbis et al. (1991, p. 319).
with a right of withdrawing goods from the market without giving anything in exchange’. As everybody knows by personal experience, nobody can finally pay by surrendering a promise to pay.

On the factor market, the payment of the ‘productive services’ involves thus three poles, that is, a firm (agent A), a worker (agent B), and a bank. The same holds for the product market, since any transaction on produced goods and services requires a seller (usually a firm), a purchaser (a household), and a bank as a ‘record keeper’ and go-between. In this regard, let us point out that gold or paper money, even in its earlier form of a goldsmith’s certificate, is just the material representation of a bank deposit, and that the transmission of bank notes or goldsmiths’ certificates between agents means, in fact, a transfer of the corresponding drawing right (that is, a financial claim) over produced output, as recorded by a bank. Every bank note corresponds as a matter of fact to a double-entry in the books of the issuing bank, a record of which the note is the material representation circulating outside the bank issuing it (see Rossi 2001, pp. 105–112). As Innes (1913, p. 407) already pointed out in this respect, ‘[a] bank note differs in no essential way from an entry in the deposit register of a bank. Just like such an entry, it is an acknowledgment of the banker’s indebtedness […]. The only difference between a deposit entry and a bank note is that the one is written in a book and the other is on a loose leaf; the one is an acknowledgment standing in the name of the depositor, the other in the name of “the bearer”’. This is a point that Courbis et al. (1991, p. 329) very clearly illustrate referring to British monetary history, particularly at the time of the first goldsmiths in London, around 1660–1665. The authors (pp. 324–325) are also clear in noting that book-entry payments existed long before bank notes or their ancestors, say a goldsmith’s certificates, appeared on earth.19

Now, in order to explain the purchasing power of a goldsmith’s certificates, or even of a gold coin (commodity money), we have to consider production and not merely exchange of already-produced goods and services. So, in the stylized example of Table 1, we notice that agent B, namely the worker, is credited with a sum (of $x$ units) of money, as written on the goldsmith’s ledger and represented by the latter certificates, for the labour services this agent delivers to what in modern parlance we dub a firm, that is, agent A. As a result, commodity money, here in the form of paper (the goldsmith’s certificates), is a means of payment that has to be distinguished from the object of this payment, which is the result of labour, that is, the worker’s effort that gives rise to

<table>
<thead>
<tr>
<th>Table 1 The result of a payment in the goldsmith’s ledger</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Goldsmith</strong> (a primitive bank)</td>
</tr>
<tr>
<td><strong>Assets</strong></td>
</tr>
<tr>
<td>Agent A (payer)</td>
</tr>
</tbody>
</table>

19. In a very interesting section of their paper, Courbis et al. (1991, pp. 329–331) note that even paper money, like bank money, is a form of credit-money. As a matter of fact, the economic foundation of any form of money is credit, not paper. In other words, paper money and bank money pertain to the same category, but of course paper money, notably in the form of bank notes, significantly increased and extended monetary circulation beyond those agents having a bank account. This, in fact, increased the size of wage-economies.
a net output for the economy as a whole. Indeed, the number of money units resulting from the payment of production costs, namely the remuneration of the worker’s effort, is the monetary form of a social product that is net for the whole economy and that defines money’s purchasing power. In a sense, the goldsmith’s certificates are worth x money units since they have been issued in the payment of factor costs for that very amount.

However, this is not the end of the story. In fact, if we push the analysis further, we can notice that, like bank money, even commodity money is not a commodity. In the contrary case, we would have to add the value of commodity money (say in the form of gold coins that the above goldsmith would have handed out to any agent claiming to obtain them in exchange for the goldsmith’s certificates) to the value of produced output, thus measuring the value of this economy’s income twice – a mistake that Smith urged us not to make.20 As a further elaboration, the rejection of the idea that commodity money is a commodity can be useful to explain the life-long, and well-known, Ricardo problem of finding a physical invariable standard of value among the set of commodities (gold, for instance).21 In short, as the social form of value, money and even commodity money must never be confused with a commodity, such as gold, paper, or any other physical item. Otherwise, we would be led to search for a non-variable standard of value among the commodity set, a problem that indeed has no solution, as Ricardo’s enduring attempt shows us forcefully.

At the time of commodity money, therefore, the money creation process was already endogenous on account of the intrinsic need for the economic system, although primitive, to express output in a social form – that is, in a form allowing for its (homogeneous) measure and exchange. To wit, commodity money is a homogeneous unit of measure since it is a dimensionless standard, an incorporeal means of payment that in itself has no value and as such cannot be measured. Referring to the above example as depicted in Table 1, we notice that the bank’s double entry (in the goldsmith’s ledger) is precisely the numerical measure of the transaction between non-bank agents A and B, because the number of (x) money units issued by the bank, or the goldsmith, measures the value of the object exchanged between A and B.

There is thus a constant circulation of debts and credits through the medium of the banker who brings them together and clears them as the debts fall due. This is the whole science of banking as it was three thousand years before Christ, and as it is to-day. It is a common error among economic writers to suppose that a bank was originally a place of safe deposit for gold and silver, which the owner could take out as he required it. The idea is wholly erroneous and can be shown to be so from the study of the ancient banks. (Innes 1913, p. 403)22

Needless to say, if the payment between A and B were carried out with a number of gold coins, instead of via the payment services provided by a bank or a goldsmith, the analysis would not change as far as the endogenous nature of money is concerned. To be sure, as noted already, the emission of gold coins had been recorded by its point of origin, namely a bank or a goldsmith, in the form of a double-entry in the issuer’s books. So, when these coins are returned to their issuer, as a deposit or in repayment of a loan, their last owner is entered on the liabilities side of the T-account kept by the issuer. Again, the law of

20. ‘In computing either the gross or the net revenue of any society, we must always, from their whole annual circulation of money and goods, deduct the whole value of the money, of which not a single farthing can ever make any part of either’ (Smith, 1776 [1978], p. 385).
21. See Ricardo (1816).
reflux shows that there can be no excess supply of commodity money, since even this primitive form of money is credit-driven and demand-determined.

To sum up, money is and has always been an endogenous phenomenon, owing to its being essentially tied to the nature of debt and the need for a means of final payment that has to be provided by a third party on the agents’ demand. Money and payments are indeed two sides of the same coin, as it were. Since any payment is elicited by the ‘needs of trade’, it follows that the money creation process is always and everywhere endogenous – even if a central bank is non-existent or non-accommodating with respect to the banks’ demand for reserves.23

5 CONCLUSION

This paper has shown that money is endogenous irrespective of its material or immaterial form. The historical evolution of the banking system has been revealing and making plain what has always been the case in actual fact: that the supply of money is demand-driven and is integrated in the real world through production, as this activity needs to be financed (as Keynes’s ‘finance motive’ pointed out) before produced output can be sold on the goods market. The evolution of banking systems provides the basis for a better understanding of Keynes’s monetary theory of production. Yet it is by understanding the endogenous nature of commodity money also that we might be able to provide some further advances in post-Keynesian monetary economics. In our view, it is indeed through a general theory of endogenous money that it will be possible to put to the fore a monetary macroeconomics analysis encompassing that of Keynes, both of the Treatise and the General Theory. The arguments made in this paper strengthen post-Keynesian monetary economics, in order to make it historically consistent, and to put it in a better position to represent a convincing alternative to traditional monetary analyses.

The position argued here is that there can never be an excess supply of money (Kaldor, 1982).24 Even in the early stages of banking (Chick 1986), the money creation process had been endogenous: neither the absence of a central bank, nor its unwillingness to accommodate the banks’ demand for central bank money (settlement balances) can be considered as evidence of money’s (partial) exogeneity. In fact, a central bank’s accommodation is instrumental in avoiding interest rate hikes that a non-accommodating behaviour would lead to in the interbank market. Behaviour, however, cannot affect an ontological principle: the nature of money is a social relationship that results from ‘the law of debt’ (Innes 1913), the material scaffold of the book-entry record being of no fundamental relevance at all for a conceptual analysis.

23. This issue goes beyond the scope of the present paper.
24. As some post-Keynesians argue, even during periods of hyperinflation price increases are not caused by excess money growth. In this respect, Camara and Vernengo (2001) argue for instance that hyperinflation is not a monetary phenomenon, but finds its origins in foreign crises. In Germany, for instance, hyperinflation during 1920–1923 was caused by the imposition of reparations and devaluation. According to Burdekin and Burkett (1992), an important feature of the German case was the way in which accelerating monetization of both government and private debt by the Reichsbank fuelled the inflation process. In fact, credit demand by the private sector arising from more rapid adjustment of money wages over that period increased fiscal influences on nominal money growth at that time. Wage claims provided the main conduit through which higher inflationary expectations were accommodated by faster rates of monetary expansion.
The conclusions that we reach here are therefore that, contrary to the argument provided by Chick (1986), money did not become endogenous over time. In fact, money has always been endogenous because of the necessarily triangular relationship involving a payer, a payee and a record keeper, even in those ancient times when money’s functions were carried out using a precious metal or, more generally, a given commodity; and this with or without the existence of ‘banks’ as such. In modern times, the banking system cannot but always respond to the needs of the economy to produce and exchange real goods and services – within as well as across borders. This is so even under a gold-standard system, as Joan Robinson (1956) noticed cogently.

If the analysis provided in this paper is correct, then the resulting conclusions will have far reaching implications both in theory and in policy. In a theoretical sense, if one can show that the money creation process has always been endogenous, then one can argue that the post-Keynesian theory of money is consistent from a historical perspective. In fact, we could argue that while money’s endogeneity actually predates the existence and development of the banking system, the visibility of this endogeneity has become more apparent in modern times. By this we mean that there is another way of interpreting Chick’s stages of banking development: the secular evolution of the banking system has made endogenous money appear more and more evident as time went by. At the time of commodity money, say in the form of gold coins, the endogenous nature of money was more difficult to notice, because money was reified into a precious metal. Hence, it would be incorrect to infer from empirical evidence, or ‘surface’ phenomena (Ingham 1996, p. 527), that money, during the gold-standard period, was exogenous because it consisted of gold bullion and/or coins, and because gold was an exogenous quantity depending on mine discoveries, wars, trade, and so on. In fact, in macroeconomics empirical evidence cannot be directly derived from factual observation: the phenomenon rarely coincides with its factual appearance. Phenomena must be interpreted and their results evaluated, which can be done only via a conceptual detour. A theoretical framework is thus always required to understand the empirical givens, such as the apparently exogenous nature of commodity or central-bank monies.

To conclude from a policy perspective, two points are worth emphasizing. First, the fact that the money creation process has always been endogenous can explain the practical difficulties and shortcomings of any monetary targeting strategies in policymaking. In fact, as repeatedly pointed out by Goodhart (1994, p. 1425), ‘[i]f the central bank tried to run a system of monetary base control, it would fail’. This also illuminates the recently generalized preference for central bankers to adopt a monetary policy strategy based on a target for inflation rather than for a growth rate of a monetary aggregate such as M0, M2, or M3 (see Rossi 2004). This change in strategy, however, is not enough to curb inflation as long as the central banks’ inflation-targeting strategies stem from a quantity-theory-of-money approach (see Rossi 2001, chs 3–4). In fact, such an approach is in open contrast with money’s endogeneity, so much so that it is grounded in the homogeneity postulate on which the neoclassical dichotomy is based, as advocated at the time of commodity money.

Secondly, the arguments developed in this paper might suggest that, rather than seeing institutional changes (the creation of a central bank and financial innovations, or even the evolution of the banking system itself) as a source of money’s endogeneity, institutional changes are the result of it. The money creation process is endogenous irrespective of financial innovations, the existence of a central bank, or the stages of banking development. In short, money is endogenous in a revolutionary, not in an evolutionary sense. This conclusion is a truly fundamental argument for post-Keynesian and heterodox monetary analyses.
As monetary and financial markets develop and evolve, the endogeneity of the money creation process implies in fact that institutions have to comply with it as well as with the double-entry bookkeeping nature of money. The framework put forward in this paper may help us, in both theory and policy, analyse the role of these institutional changes properly.

REFERENCES


