Introduction

Non aes, sed fides

If the success of an institution can be fairly judged by its diffusion, then the central bank is without doubt a very successful institution. When Will Rogers, the American humorist, said, ‘There have been three great inventions since the beginning of time: fire, the wheel, and central banking’, it may have been mere journalistic hyperbole, but like all hyperboles it contained an element of truth. In 1900 there were only 18 central banks. By 1950 that number had risen to 59 and by 1999, the last year for which accurate figures are available, it was 172. Today, over 90 per cent of the United Nations member states have a central bank.

More telling than these figures, which reflect the proliferation of nation-states following, first, the break-up of the great European empires and then the demise of colonialism, is the change that took place during that period in the perception of the role and the prerogatives of the central bank and its place in the organization and apparatus of the state. At the turn of the twentieth century, the central bank, still known by the less imposing name of ‘issuing bank’, was an institution with an uncertain future. Not all of the leading countries had one, notably the United States, which had a long-standing distrust of centralized power. In other countries, such as Italy and France, the central bank continued to lead a precarious existence between balance sheets constricted by large volumes of non-liquid financial assets, trifling legal constraints, and constant debates on the advantages of decentralizing issuance. Nowadays, the central bank has become an essential attribute of the state, if not a symbol of sovereignty.

Given the importance attributed, for better or for worse, to the central bank it may seem surprising that until quite recently it attracted very little theoretical interest. As Fausto Vicarelli remarked some 15 years ago on the question of the central bank’s ‘independence’ from political power:

When the economist begins to tackle such an important and topical subject as central bank autonomy, one question immediately arises: can the autonomy or independence of an economic institution be discussed without taking into consideration a theory which justifies its existence and explains the logic of its evolution? Assuming there are grounds for this question, if the answer to it is a negative one, the economist who begins to explore central bank autonomy will
encounter certain difficulties since, even today, central banking theory has yet to be developed or, more optimistically, is still in the development phase.²

There are many reasons why it has taken so long for a theory of the ‘evolution’ of central banking to emerge. The main reason is probably entirely analytical, that is, that the neo-classical theoretical paradigm’s great heuristic potential in other fields does not reach to institutions – any institution, not just the central bank. In attempting to create a more rigorous discipline, neo-classical scholars have restricted both their method and their field of analysis to the point of reducing their theory of institutions virtually to an oxymoron. Firms, trade unions and even money are thus left out of the neo-classical paradigm.

Without a coherent theoretical framework, it was easy, given the close relationship that had linked central bank and state since the former began to evolve, for the adversaries of central banking to attribute the institution’s existence to the prince’s greed, to the constant attempts of the person in power to entangle the economy in a web of privileges and prohibitions designed to increase the power to raise taxes, whether openly or secretly. We can call this the ‘fiscal’ theory of central banking; Friedrich von Hayek and James Buchanan were its founders, Kevin Dowd, Lawrence White and David Glasner are its present leading exponents.³

The fiscal theory of central banking is backed by an incontrovertible historical fact, that in the early days of many central banks the state’s action was specifically designed to reap advantages from the creation of a large bank endowed with special privileges enabling it to mobilize rapidly substantial amounts of funds with which to finance war. This passage in institutional history was an experience shared, for example, by the Bank of England, the Banque de France, the Bank of Prussia and the Banca Nazionale nel Regno d’Italia. Nonetheless, it is anachronistic, and often blatantly incorrect, to argue on this basis that such a mechanism is the raison d’être of central banks as we know them today. Other central banks, for instance, were not created for an immediate return in the form of financing public expenditure but for the advantages to the community of rationalizing the payment system, which was in complete chaos at the time, whether by historical accident or from poor legislation. This was the case of the Reichsbank, the Banca d’Italia, the Schweizerische Nationalbank and numerous other smaller institutions. Moreover, the nature of the financial advantage that would accrue to the state from setting up a privileged bank is unclear. For instance, on closer inspection, the creation of the Bank of England can be attributed to an attempt on the part of Britain’s emerging middle class to exert fiscal control over the sovereign, not to facilitate secret taxation. The state certainly gained in
creditworthiness. All told, identifying the contingent factors that lead to the creation of this or that issuing bank is not the same as searching for the evolutionary mechanism that produced what we now term a central bank. For that, fiscal theory is of little help.

At the time that Vicarelli voiced his disappointment about the state of the theory of central banking, Charles Goodhart exploited the intervening progress in the analysis of the informational problems of financial markets in a book that marked a major step forward in the thinking on this subject. Goodhart viewed the central bank as the fruit of an institutionalization of restrictive competitive practices required by the very nature of banking. He argued that because bank loans cannot be traded on secondary markets, depositors find it difficult to assess their true value. Hence, doubts about the soundness of a bank’s capital can spread to the whole system, placing its stability at risk. It is therefore in the interests of the soundest banks to club together in order to safeguard their reputation, adopting any method of selection and rule of conduct for their members that will minimize opportunistic or fraudulent practices. However, and this is Goodhart’s main thesis, such a governance structure is unlikely to emerge spontaneously or to be able to withstand the clash of interests that would ensue from increased competition in the credit markets. For the club president to have sufficient authority to issue rules, ensure their observance, and impose sanctions on recalcitrant members, he must remain outside the fray and not enter into competition with the subjects he governs. This would explain the state’s intervention in the banking sector. By legislating to support a club structure hinged on a non-competitive central bank that is not constrained to maximize profit, the state exercises its coercive powers for the purpose of maintaining internal discipline. This is not an arbitrary action, therefore, but one designed to safeguard the stability of the banking system and so, in the final analysis, also the interests of depositors and of the economic system as a whole. All in all, such a development is a ‘natural’ one, as Goodhart stated in the subtitle to an early edition of his essay, published in 1985.

Goodhart’s analysis has several merits, including the fact that it is presented as a logical extension of the doctrines of the ‘fathers’ of Anglo-Saxon-style central banking such as Henry Thornton and Walter Bagehot. It also has some evident limitations. To begin with, it is not a ‘monetary’ theory of central banking. If the problem of banking crises related only to the opaqueness of loans, it would be sufficient to separate credit circuit from money circuit in order to root out instability, following an old prescription dear to the monetarist school of thought. Why does a bank account that can be used for payments seem so inextricably linked to the granting of non-negotiable loans? What has led modern societies to accept
such an evidently explosive mix and then ‘invent’ the central bank to
defuse it? Moreover, in Goodhart’s vision the central bank is born already
fully developed, as it were. Against all evidence it undergoes no evolution,
either in its functions or its instruments, and even less in its institutional
role. On top of this, Goodhart’s central bank has nothing to do with
monetary policy or with the payment system, two areas that are instead
in the forefront of discussions among central bankers and of reconstruc-
tions of the history of the individual central banks. Last of all, how can
one explain the transformation of a decentralized banking system into a
pyramid structure with the central bank at the top? Studies by Mancur
Olson and his followers have familiarized us with the notion that it takes
more than a shared interest for a group of individuals, however small, to
decide to cooperate together. Remaining on the subject of banking, how is
it possible that where a tight-knit group of bankers failed, politics, which
depends on the consensus of much broader sections of the population, can
succeed, not forgetting that the advent of central banks pre-dates univer-
sal suffrage? And how can one explain the repeated connection between
banking reforms and crises of confidence? Or the fact that the evolution
of central banking has been interspersed with ‘waves’ of reform affecting
numerous countries whose legislators appear to have been inspired by the
same model, adapted in each case to local conditions?

None of these questions are answered in Goodhart’s analysis. Indeed,
they are never even raised. The reasons for this silence are easily under-
stood and come down to two. First, there is the aim of reflections at
the time, which was essentially polemical. Goodhart’s contribution was
intended as a response to the reappearance in academic debate of what I
have called the ‘fiscal’ theory of central banking. During the proliferation
of new labels such as the ‘theory of legal restrictions’ or ‘new monetary
economics’, all sharing a profound aversion for state intervention in
monetary matters, it was a way of drawing the attention of academia and
politics back to the ‘lofty’ motives for giving substance, a century earlier,
to the embryonic institution of the central bank. Among such ‘lofty’
motives, the performance of the role of lender of last resort in financial
emergencies was undoubtedly the most important. Second, Goodhart’s
analysis reflects the tools available to him at the time, which, despite the
addition of the concept of imperfect information, barely acknowledged
the concepts of transaction cost and uncertainty in the Keynesian sense.
As Oliver Williamson demonstrates in his analysis of firms, Avner Greif,
Barry Weingast and Paul Milgrom in their studies of mediaeval corpora-
tions, and Douglass North and Barry Weingast in relation to the English
revolutions of the seventeenth century, it is virtually impossible to describe
complex institutional changes without somehow linking the underlying
assumptions of theoretical analysis to the reality being explained. In the rarefied world in which substantial rationality, stable preferences and parametric uncertainty coexist, no role can be found for institutions.

Therefore, while Goodhart’s important contribution must be the point of departure, it must be taken further by building around it a complex mosaic of many hues and shades. Lawrence Broz attempted this in 1997 in an essay on the creation of the Federal Reserve System. He specifically followed the route carved by Goodhart but proposed his own solution to the problem of explaining the institutionalization of central banking, that is, the establishment of a body for the pursuit of a public good, financial stability, in a world of atomistic agents dedicated to maximizing their own personal utility function. That solution is based on the concept of joint production: in the case of the United States, the central banking model became acceptable when a small group of bankers at the head of the leading New York banks were finally convinced of the advantages centralization would bring in terms of greater international use of the dollar. By promoting legislation in favour of central banking, and hence the public good of financial stability, at their own expense, those bankers in effect reaped benefits from the spread of their business across the Atlantic as the dollar became a currency of international reserves and trade.

The theory of public and private joint production may help to explain the sensational about-turn that took place in the United States, a country with a deep and long-standing aversion to central banking, with the passage of the Federal Reserve Act of 1913. However, it still does not explain why the about-turn occurred when it did and not earlier or later. Indeed, the radical nature of that institutional innovation begs the question how a group of bankers, however powerful, could successfully oppose the rest of the nation, and even suborn it to their own ends. However, it is when Broz attempts to apply his explanatory model to other countries that he becomes less convincing. In order to keep faith with the notion of joint production Broz is forced to find a private commodity with broad enough significance to elicit demand for reform even in countries with no ambition for an international role of their currency. That commodity, once again, is ‘fiscal’, although Broz takes account here of the lessons of North and Weingast on the origins of the Bank of England: the fiscal mechanism in this case is the benefit to the state from an improvement in creditworthiness following the creation of a central bank.

The explanation is not convincing for a number of reasons. While, as noted earlier, considerations of a fiscal nature do indeed lie behind the evolution of many central banks, equally it is anachronistic to try, at that early stage, to attribute the emerging bank with the role of guardian of financial stability, which only developed much later. The Bank of England
of 1694, for instance, is a far cry from the Bank of England of 1913, and resembles even less the Federal Reserve set up in that year. Moreover, for some central banks such as the Banque de France and the Bank of Prussia the fiscal objective was not so much to improve the state’s creditworthiness as to provide the Treasury with low-cost financing. The fact is that trying to reduce the long and varied evolution of central banks to a single mechanism – one adhering, moreover, to the assumptions of methodological individualism and substantial rationality – entirely out of regard for the theoretical clarity of an analytical apparatus, that of neo-classicism, evidently unable to offer tools of interpretation from an institutionalist viewpoint, is not a laudable enterprise, either on the theoretical level or, even less, on the historiographical one.

These are the premises behind the attempt that resulted in this book. The idea from which the analysis stems is that in order to understand the central bank, it is necessary first to understand money. The mistake that neo-classical theories make is to consider money a commodity. Instead, it is really an institution that is held up by trust: trust in its future purchasing power and trust in the continued convention that payment is complete when money changes hands (even if this is now something of a metaphor). The institutional nature of money is emphasized by building the analysis around the concept of ‘payment technology’, meaning the set of conventions, objects and procedures that allow obligations arising from trade to be extinguished. The fundamental problem of any payment technology is the incomplete monetary contract: it is impossible to define the future value of money beforehand without socializing the whole economy. Owing to this uncertainty, the spread of a given payment technology will depend on the existence of institutions and roles to safeguard users’ trust in it. This is because as soon as a stock of money has been built up, the resulting utility flow becomes a quasi-rent, which can be expropriated by manipulating the supply of money. From this standpoint, the evolution of central banking is merely one aspect of the institutional adaptation set in motion by the development of payment technologies; as these became increasingly abstract and hence easier to manipulate, they made it more and more difficult to preserve the value of the outstanding stocks of money over a period of time.

In this approach the problem of transition from one payment technology to another inevitably becomes important. Once a particular payment technology has taken hold, why does the community feel the need to develop another, and how does the process of innovation occur? Who are the participants, what are the constraints, and what are the relative roles of the market and of politics? In order to shed light on the problem of transition we need to start from another, often neglected, characteristic
of money: that it is present as counterparty in all exchanges taking place in the economic circuit. This characteristic, from which it follows that the price of money is none other than the inverse of the average of the prices of all other goods, means that any imbalance in the money market, given rigid supply, can only be eliminated by making adjustments in all the other markets. Nowadays, we imagine a typical money market shock as a sudden, unexpected increase in supply, because our way of looking at monetary matters is shaped by our familiarity with a form of money, token money, that can be manipulated at pleasure. However, for much of recorded history the main problem was the very opposite. From the late Middle Ages on, financial innovation was driven by the need to find more flexible forms of money in order to counter the long-standing tendency of the demand for money to increase as a result of economic development and the growing social division of labour. Thus, monetary innovation arises out of the need to counter the inherent deflationary drift of payment technologies based on rigid money supply, at least in the short to medium term. The fact that the process did not immediately lead to the generalized adoption of a token money was due not so much to technical limits – paper credit money was well known to Renaissance merchant bankers and paper money was already widely used in China in the Middle Ages – as to a lack of institutions to uphold confidence in payment technologies built around a money that had no intrinsic value.

With the industrial revolution and virtually contemporaneous development of the representative state a structural split occurred. On the one side, as the economic circuit became increasingly complex it fuelled the social incentive to develop more flexible payment procedures. On the other side, under the new political and institutional framework monetary institutions could, for the first time, develop outside the control of the prince. Any attempt to move beyond commodity money, even in its most advanced form of coinage, must entail an intermingling of money circuit and credit circuit. This is the great innovation implicit in the notion of banking, in a process wonderfully described by John Hicks:

This is the point at which deposits in banks, withdrawable deposits, are made transferable: either by cheque, which is an instruction to a bank to transfer an existing deposit, or by note – which is in effect a cheque payable to bearer, having the guarantee of the bank behind it, without reference to the depositor against whose deposit it was originally issued. This is vital; for it is at this point that the bank becomes able to create what is in effect money. When it makes a loan, it does not have to hand out the old ‘hard’ money; all it does is to exchange claims. Against the obligation of the borrower, to repay by some fixed date, it provides an obligation of its own, which is transferable upon demand, and for that reason has a money quality. The money which it lends is money that it itself creates.
The intermingling of money and credit thus set in motion a long and somewhat tortuous process of institutional adaptation centred around the figure of the central bank. To date, the process has consisted in three separate phases. In the first phase, believed to have ended in 1844 with the adoption of the Bank Charter Act in England, which was imitated and adapted throughout the world in the following decades, it became an established principle that the issuance of convertible banknotes was a 'special' activity, subject to limits and performed under the control of the state. This period saw the emergence of the first great issuing institutes operating under a monopoly, although in several countries there were still forms of monetary competition, albeit subject to tight restrictions. In the second phase, the problem was how to govern bank money, which developed as banknote supply became increasingly rigid during the earlier round of reforms. This was when the private issuing institutes were transformed into true publicly owned or semi-publicly owned central banks. Operationally, it was also during this phase that the central banks became lenders of last resort and banking supervisory bodies, although how and when this happened varied from one country to another. Finally, the third phase was marked by the introduction of the inconvertible banknote as legal tender and of the concept of 'managed currency' – the revolutionary idea that monetary policy should actively aim to boost production and increase employment. Institutionally, this phase saw the definitive nationalization of the central banks and subsequent rise of the ‘principle of autonomy’ announced by David Ricardo in an essay published posthumously in 1824. According to Ricardo a condition for the successful control of the money supply by the state is that the technical body responsible for monetary policy should not be subject to the power of the executive but should answer to the legislature for its actions after the fact. At the same time as the notion of managed currency took hold, new methods of operation in the conduct of monetary policy were developed, notably open-market operations.

The three phases described above are defined on the basis of logic rather than of time. They occurred at different moments in different countries and not all of them in every country. Their contours are often hazy. For example, when it came to the point of regulating convertible banknotes, the spread of bank accounts had already begun. Similarly, the wave of legislation of the 1920s and 1930s, marking the final acceptance of payment technologies centred around the bank account, occurred when experimentation of the fiat standard had been under way for some time. Probably, we have now entered a fourth phase, in which electronic systems, and the consequent possibility of intra-day credit, determine the how and when of
a new wave of institutional adjustments, spilling over national borders for the first time.

The neo-institutionalist approach has several implications that should be highlighted. First, monetary institutions are viewed as the result of a continuous adaptation, the form of which is dictated by contingent problems and by the existing political and institutional heritage. A crucial role in this process is played by the concept of ‘crisis’. A crisis occurs when an innovative payment method shows itself to be precarious and a demand for institutional reform takes shape. The link between crisis of confidence and reform process is a macroscopic phenomenon of monetary history. Yet economists, being theoreticians, have accorded it only passing attention, primarily to minimize its real practical relevance. Instead, this essay takes the view that a crisis is an inevitable step in the process whereby society learns the potential and limitations of a given payment technology.

A demand for reform is not necessarily matched by a supply of reform. This is because the concept of supply of reform involves the political sphere. Thus not only must the demand for reform have reached such proportions as to engage the political sphere, but there must also be a class of political entrepreneurs able to promote and formulate an institutional answer and gather sufficient consensus around it. It is at this point that the problem of collective action comes into play, which Broz studied in relation to the United States. My theory in this regard is that producers of an innovative payment technology play a major role in shaping political supply and need the quasi-rents associated with their investments to be safeguarded in times of instability. In other words, in no way does regulation tend to punish spontaneous market innovation, indeed it is often a way of reducing the likelihood that the producers will behave opportunistically. This is why in most cases it leads to a decrease in competitive behaviour and to the introduction of entry barriers in the monetary sector of the economy.

The important criterion for establishing the point where demand and supply of reform meet is not and should not be the criterion of efficiency, but that of effectiveness. In economic language the word ‘efficiency’ is often used randomly, as if it expressed an unambiguous concept. While this way of thinking may cause little damage in other fields, in the analysis of institutional processes it can be dangerous. Every institution is an answer to a problem of incomplete contract caused by Keynesian uncertainty or by prohibitive transaction costs. Given this, judging efficiency with respect to an ideal that does not exist is not an approximation, it is a serious logical error. Harold Demsetz has called the tendency to judge an actual situation with reference to an idealized situation that cannot happen the ‘nirvana fallacy’:
The view that now pervades much public policy economics implicitly presents the relevant choice as between an ideal norm and an existing ‘imperfect’ institutional arrangement. This *nirvana* approach differs considerably from the *comparative institution* approach in which the relevant choice is between alternative real institutional arrangements.¹⁰

A more useful notion of efficiency, according to Demsetz, would refer to situations of scarcity and to individuals’ knowledge and preferences as they really are, not as they should be. For instance, it is only meaningful to refer to a suboptimal state if it is possible to achieve a situation that is effectively better than the present one. However, as we know, there are no theoretical instruments beyond Pareto improvement that allow such a comparison to be made in abstract terms. In other words, there is no way of evaluating the ‘naturalness’ of a monetary system or its efficiency. In the monetary field, too, the only proof of the pudding is in the eating. That is why this study never refers to the concept of efficiency, preferring that of *effectiveness* in respect of the problem perceived to be socially significant at a given point in history and in the light of existing information and institutions. It is because of the structural uncertainty of the outcomes that monetary reforms tend to operate on the fringe of existing systems without overturning them. The history of monetary institutions offers striking examples of what the neo-institutionalists call ‘path-dependence’, the tendency of present choices to reflect those of the past.

This consideration leads to another of a more general nature. Putting the concept of effectiveness first means broadening the analysis beyond the strictly economic sphere because the feasibility of a given institutional situation depends on the legal and political order – in neo-institutionalist jargon, on the ‘institutional environment’. Many examples of this can be offered but here they are confined to two. The spread of banknotes, which sparked the evolution of central banking, would not have been possible if the legal system had not accepted the concept of the bearer security, granting it the same degree of protection as the registered security. It was because of the lack of a suitable legal superstructure that payment technologies with a credit content (already circulating widely among merchants since the fourteenth century thanks to the protection afforded by customary mercantile law) did not become current among large sections of the population until well into the eighteenth century. It would be hard to understand how the concept of legal tender could gain worldwide acceptance without considering the intervening change in the role of parliaments, from bodies overseeing the deeds of the prince to depositaries of popular sovereignty expressed by universal suffrage. The examples abound.

Finally, by stressing effectiveness over efficiency it is possible to explain why the necessary phases in the evolution of central banking were marked...
by waves of reform, leading to the reproduction of the same institutional model in a multitude of countries, with lesser or greater adjustments according to the local context. This phenomenon would be impossible to explain otherwise, as pointed out earlier in connection with Broz’s interpretation of the institution of the Federal Reserve.

At this point, having outlined the essential features of the analysis that follows, it would be appropriate to define the aim and limits of the present exercise. The book does not have historiographical ambitions because its author has neither the training nor the interests of a historian. Most of the sources quoted are secondary sources and the history of central banking is not examined in its entirety — spatial, temporal or functional — but only insofar as it affects the illustration of theories. If a label has to be found for the present approach, then a suitable one would be ‘theory of history’, of which an authoritative example can be found in some of the works of John Hicks and which has been revived on several occasions even in recent years.11

A theory of history, the boundary line suggested by Hicks, deals with general phenomena not specific events and it is obliged to use abstractions, with which specialist historians are often unable to identify. Its usefulness, however, should be judged according to whether it is able to shed light on phenomena that recur in different places and times, and not according to its accuracy of description. As Hicks himself pointed out, however, the main risk of a theory of history is that it may be guilty of determinism. Once the mechanism for reproducing comparable institutions or events has been identified, it is easy to succumb to the temptation to view its operation as inevitable. Although this would be seriously wrong for any historical phenomenon, it is particularly dangerous in the case of the history of money. The ‘technology’ embodied in a payment technology is largely social because of the role played by trust: therefore, it is evident that its diffusion and the events affecting it, including the fact that it might degenerate, reflect the larger vicissitudes of the community adopting it. If there is ‘progress’ in the monetary field, it is progress in the sense of ‘progressive’ affirmation on a vast scale, from the Industrial Revolution on, of payment technologies with a highly flexible supply and that are consequently easier to manipulate. Whether they are then manipulated in the collective interest or for private interests, or in the long term maybe in the interest of no one, will depend on the soundness of the institutional environment and its juridical and political components.

Put differently, the history of money and its management is anything but linear. In order to emphasize how the social aspect takes precedence over the merely technical side, Chapter 2 discusses the credit payment technologies in use in Europe long before the industrial revolution, and
which disappeared once the economic and institutional bases for their use no longer existed. From this standpoint, current payment technologies owe their progress to the institutions of the liberal and democratic state, founded on concepts of the division and limitation of power. Nominalism aside, the central bank and the liberal and democratic state are joined in a truly symbiotic relationship: the same origins, the same development, and in all likelihood the same future, whatever it may be. Without liberal and democratic institutions, capitalism would probably be unthinkable, a central bank certainly so. That is why Keynes’s dream of a world central bank collapsed at Bretton Woods. It is also why in Europe in the 1990s the creation of a central bank that transcended national borders was only possible because it was part of a much broader project to institute a federal political structure.

Benjamin Cohen may be right that the geographical future of money will see an increasingly clear dissociation between the state and the institutions responsible for the management of that money. However, it is still too early to say whether this is an established tendency or just a stage in the process of redefining the scope and structure of the multinational state. It is present history and much of it has yet to be written. The next stage in the evolution of central banking will very likely depend on its outcomes and its teachings.