The foundations of money, payments and central banking: A review essay

Stephen Millard

Bank of England, Threadneedle Street, London, EC2R 8AH.
E-mail: stephen.millard@bankofengland.co.uk

Contents

1 Introduction and motivation

The purpose of this review essay is to understand the economics behind the evolution of payments. By payments I mean the 'transfer of monetary value' (in return for goods, services, or real or financial assets). Understanding the evolution of payments is important from the point of view of a central bank since they typically take a major role in payments: in the provision of a settlement asset for large-value payment systems, often in the ownership and operation of such systems and also overseeing these and smaller-value, retail payment systems. In deciding what, in fact, should be the core roles of central banks in payments, it is important to understand how central banks are linked to payments and how they jointly evolved.

One way of answering the question as to what roles a central bank should take on in payments is to start from their core purposes – the provision of monetary and financial stability – and ask what roles these imply. This essay takes a different approach. It asks why an economy would benefit from the presence of a central bank in the first place and assesses the implications of the answer to this question for the role of a central bank in payments. As an aside, taking this approach also allows us to show why central banks should be concerned about monetary and financial stability. In other words, this essay argues that central banks developed to perform a payments role and that their core purposes of monetary and financial stability stemmed from their performing this role.

It is clear from the definition of payments given above that, in order for there to be payments, there first needs to be money. In particular, in a barter economy where goods are exchanged for goods there are no 'payments' in the sense that I am using the word. So, in Section 2, I discuss why money might evolve as a result of some frictions inherent in real-world economies. Section 3 then discusses the evolution of banks. It argues that banks developed in order to provide payment services (making 'money' work more efficiently); banks as intermediaries channelling savings into investment came later. Section 4 discusses how banks can save on the use of collateral to make payments – collateral that they can convert into loans to earn a return – by the development of 'payment systems'. Such systems will involve some form of netting of payments (clearing) and final settlement in some asset. 'Central banks' fit into this picture by providing, in their liabilities, a settlement asset that the other banks are happy to use. In so doing, they are incentivised to worry about monetary and financial stability. Section 5 concludes with some (tentative) thoughts on the issue of a central banks role in payments today.

2 Why have money?

Money is commonly defined in economics textbooks with reference to its functions. In particular anything that is generally acceptable as a medium of exchange, unit of account, store of value and means of deferred payment. For instance, Baumol and Blinder (1991) define money as "the standard object used in exchanging goods and services. In short money is the medium of exchange." They go on to say, "But once it has become accepted as the medium of exchange, whatever object is serving as money … will inevitably become the unit of account … Money may also come to be used as a store of value."

In terms of the economics of payments, the key aspects of this definition are 'medium of exchange' and 'means of deferred payment'. Indeed, it is interesting, at this point, to pose the question as to whether or not the unit of account need be related to the medium of exchange. Even in a pure barter economy with no medium of exchange and n goods, say, having a standard unit of account would be useful since it means that agents would not have to know all possible relative prices in the economy $\left(\frac{n(n-1)}{2}\right)$ but simply the price of each good relative to the unit of account (n-1 prices).

In fact, as pointed out by Rosenblat (1999) there are examples of barter societies that had units of account. She cites Grierson (1977), for example, who, in turn cites documentary evidence from ancient Egypt, c. 1275 BC, in which a slave girl worth about 373g of silver was traded for clothes, blankets and a miscellany of other objects worth the same amount. He also provides a couple of examples from the *Iliad* and the *Odyssey*. In one example, suits of armour were exchanged, and in the other example, gold was exchanged for a slave, but the prices of all of these were expressed in terms of oxen.

More recently, there were many examples in late mediaeval Europe where monetary systems were set up with a standard coin as the unit of account but where the coin disappeared from circulation long before it stopped being used as the unit of account; that is, the unit of account differed from the medium of exchange. Two examples cited in Quinn and Roberds (2005) are the French ecu of 1577 which operated as a unit of account for 25 years after it ceased circulating and the 1543 Dutch silver florin which had vanished from circulation by 1609 (when the Bank of Amsterdam was founded) but was still the unit of account in most of the Dutch Republic at that time.

But why do we need a 'medium of exchange'? In the Arrow-Debreu world of general equilibrium and complete markets, we would not need to have a medium of exchange. Indeed, microeconomic textbooks will typically consist of models in which there is no money and even macroeconomic models generally assume the existence of money rather than showing that it will exist in equilibrium. So what is the imperfection that creates the need for money? The model of Kiyotaki and Wright (1993) provides an answer to this question. In what follows, I follow the description of this model given in Rupert *et al* (2000).

Consider an economy with a unit continuum of infinitely-lived agents with discount rate r. Each agent has the ability to produce one unit of a non-storable specialised good at a cost, c. There is also a unit continuum of such specialised goods. Agents do not like all goods, only some. More concretely, I assume that a particular agent, i, will derive utility, u, from consuming a good produced by a different agent, j, with probability x and will derive no utility from it with probability (1-x). To ensure that there are gains from trade, I assume that $u > c$. I assume that the conditional probability that j likes consumer i's good conditional on i liking j's good is y. So the probability of a 'double coincidence of wants' is xy. Intuitively, it will be the fact that this probability is 'small' that leads to a role for money in the economy. Finally, I assume that any agent i derives no utility from consuming his own good. In this economy, there are also M indivisible units of a storable good – 'money' – that is intrinsically worthless, in that consuming

it generates no utility. At the beginning of time, a random group of *M* agents are each endowed with one unit of money. I assume that no one holding money is able to produce a good.

Agents trade bilaterally in this economy; there is no central market where they can all meet. Instead, they meet each other according to a random pairwise matching process with Poisson arrival rate equal to unity. I assume that the history of all agents' past meetings and trades is private information and cannot be found out by anyone else. Suppose agent *i* meets agent *j*. If there is a double coincidence of wants, then they will trade with each other. Alternatively, if *i* is carrying money and *j* can produce a good that *i* values, then *i* will wish to trade. Whether or not *j* would be happy to trade with *i* in this case will depend on whether or not he believes that other agents will accept money from him in the future. Suppose that *j* believes that a random agent will accept money from him in the future with probability β. I consider stationary and symmetric equilibria in which the probability that *j* will accept money from *i* is also equal to β. Ican write the value functions for those agents carrying money (V_1) and those agents not carrying money (V_0) as follows:

$$rV_1 = \beta x(1-M)(u+V_0-V_1) \quad \textbf{(1)}$$

$$rV_0 = xy(1-M)(u-c)+Mx\beta(V_1-V_0-c) \quad \textbf{(2)}$$

Now, there will always exist at least one equilibrium: that in which no agent accepts money, i.e., $\beta = 0$. This demonstrates how the existence of (fiat) money depends on people believing that enough other people will accept it. Note also that, if a monetary equilibrium exists, there will be two symmetric equilibria: that in which $\beta = 1$ and another in which β equals some value between zero and one.

Denote the net gain from trading goods for money by $c_0 = V_1 - V_0 - c$ and the net gain from trading money for goods by $c_1 = u + V_0 - V_1$. Solving for these implies:

$$c_0 = \frac{x(1-M)(\beta-y)(u-c)-rc}{r+\beta x} \quad \textbf{(3)}$$

$$c_1 = \frac{x(\beta M+(1-M)y)(u-c)+ru}{r+\beta x} \quad \textbf{(4)}$$

As I said earlier, a non-monetary equilibrium will always exist but, for the two monetary equilibria to exist, we need $c_0 \geq 0$ and $c_1 \geq 0$. Clearly the second condition holds for all possible values of our parameters.

Monetary equilibria will exist provided agents are sufficiently patient since agents pay the cost of producing today to obtain money but only benefit from carrying money in the future when they meet someone who can produce a good they like and who will accept money. Setting β equal to one and using the condition that $c_0 \geq 0$ implies:

$$r \leq \frac{x(1-M)(1-y)(u-c)}{c} \quad \textbf{(5)}$$

Similarly, monetary equilibria will exist if the gains from trade are sufficiently high. Adopting a similar approach yields the two conditions:

$$u \geq c + \frac{rc}{x(1-M)(1-y)} \quad \textbf{(6)}$$

$$c \leq \frac{ux(1-M)(1-y)}{x(1-M)(1-y)+r} \quad \textbf{(7)}$$

So, I have shown that, given certain parameter restrictions, monetary equilibria exist in this model; in other words, there is a potential role for money in eliminating the 'double coincidence of wants' problem. But is the introduction of money welfare improving? To examine this define welfare, *W*, by:

$$W = (1-M)V_0 + MV_1 \quad \textbf{(8)}$$

In a pure barter economy, $rW = xy(u-c)$. In a monetary economy, $rW = x(1-M)(u-c)\big((1-M)y + M\beta\big)$. Given this, we can see that the introduction of money into the economy improves welfare provided:

$$\beta > \frac{y(2-M)}{1-M} \quad \textbf{(9)}$$

In other words, for the introduction of a given amount of money into the economy, this will be welfare improving provided enough agents accept it. Setting β equal to one implies:

$$M < \frac{1-2y}{1-y} \quad \textbf{(10)}$$

In words, this condition tells us that the introduction of money into the economy only improves welfare if there is a sufficiently small probability of a double coincidence of wants.

But is this the only friction necessary to give money a role in the economy? Kocherlakota (1998) asked whether there were other ways of overcoming the double coincidence of wants problem without using money. He showed that, for money to be necessary in this sense, two other frictions were needed in the economy: agents being unable to commit to repaying loans and trading histories of agents being private information.

To see the importance of commitment consider what would happen if all agents adopted the strategy of producing for anyone who wanted their good. If all agents could commit to this strategy, then there would be no need for money. But even in a world in which agents cannot perfectly commit, it is still possible for this strategy to be an equilibrium if all actions are public information. (Note that the strategy would need to include a specification as to what would happen once someone deviated; a simple strategy would be to revert to only producing when a double coincidence of wants occurred.)

Kocherlakota (1998) formalised this in the following way. He assumed that all transactions were recorded on a spreadsheet and that before agreeing to a trade an agent can costlessly observe the entries in this spreadsheet that correspond to the agent he is about to trade with, all agents that they have traded with in the past, all agents that these agents have traded with in the past, and so on. So, if you went into a shop and asked for coffee, say, the shop-keeper could access the spreadsheet and check that you had produced enough for other people in the past to warrant being given coffee now. If you had, he would simply give you the coffee. Your 'account' on the spreadsheet would be adjusted accordingly. Of course, the spreadsheet doesn't exist in practice and the shop-keeper is unable to assess whether you have produced your good for other agents in the past, i.e., your trading history. In its place, money acts as the system's memory.

Another way of thinking about this problem is to note that the equilibrium of all agents adopted the strategy of producing for anyone who wanted their good can be supported through agents taking actions that make it costly for them subsequently to deviate. One such mechanism is the use of collateral – that is 'buyers' assign control over some assets to 'sellers', gaining control over other assets once they provide a service to someone else. If an agent deviates, he would lose his collateral for ever; this leaves a strong incentive not to deviate. But, in this system buyers are simply exchanging assets for goods. Hence, 'money' can be thought of as simply a replacement for collateral: a 'collateral asset'.

So, I have shown that welfare is improved by introducing money into an economy with a 'double coincidence of wants' problem, limited commitment and private trading histories. However, I have not discussed what this 'money' actually is. The model discussed above has been set up as if money were purely fiat money, though it can easily be extended to the case of commodity money (as done in, say, Velde *et al* (1999)). Indeed, these represent two different models of the emergence of money in practice.

One model involves a government printing pieces of paper – this can be done either directly or by a central bank (an issue to which I will return) – and forcing the acceptance of these. Typically, they could do this by making the notes 'legal tender' (meaning that creditors have to accept them from their debtors) or by forcing taxes to be paid using them, since agents would then know that they had a use for any notes they accepted from other agents. One particularly brutal example of this approach can be found in the Mongol empire in the late thirteenth century described by Marco Polo in his 'Travels in the land of Kubilai Khan'. Kubilai Khan had paper currency printed and distributed it by 'buying' gold and jewels from his subjects using the notes. He then imposed the notes as legal tender and imposed the death penalty on anyone who did not accept them. Another example can be found in 19th century Canada where 'Dominion Notes', printed by the Government and only partially backed by gold, became legal tender and also the settlement asset for interbank payments.

The alternative model involves a gradual transition from commodity money to fiat money as private individuals (bankers) issued pieces of paper carrying a promise that they could be redeemed in terms of the commodity money. As these notes circulated more and more and came to be redeemed less and less, they became 'money' and, as banks started lending, the quantity of

notes in circulation came to be less related to the amount of commodity money backing them. In the following section, I describe this story in more detail by considering how the need to safely store money led to the emergence of deposit banks and, as a corollary, the need for an interbank payment system.

3 Why have banks?

In the previous section I examined how a collateral asset called 'money' – that could take the form of either government-issued pieces of paper or a commodity with some intrinsic value – can solve the problems associated with a lack of any double coincidence of wants, lack of commitment and private information about trading histories. The question for this section is: Can an economy develop arrangements involving privately-issued debt (inside money) that enable it to get closer to the Arrow-Debreu equilibrium in the face of a limited supply of this collateral asset? Such arrangements are what we have come to think of as a 'payment system'. Indeed, the Bank for International Settlements defines a payment system as 'a set of instruments, banking procedures and, typically, interbank funds transfer systems that ensure the circulation of money.'

Of course, before there could be a set of 'banking procedures' there needed to be 'banks'. One story for the emergence of banks is that they were originally moneychangers. Using commodity money (e.g., silver coins) to make large-value payments was inefficient given the relatively low values of the coins and the wide variety and quality levels of different coins circulating; it was much more efficient for merchants to deposit their coins with moneychangers – who had a comparative advantage in being able to value coins – and for transactions between merchants to happen across the books of the moneychangers rather than for coins to be counted out and valued for each and every transaction. In this way, moneychangers became early deposit banks. (See Kohn (1999).) Payments were made with both payer and payee present at the bank. Where the payee did not hold an account at the payer's bank, he either opened one or could ask the bank to transfer the money to his own bank. Since banks were close to each other, this was done by the payer's banker walking over to the payee's banker with the money.

Another reason why merchants would deposit their coins with moneychangers, gold and silversmiths and even innkeepers and monasteries was for safekeeping: carrying money around is unsafe since it can be stolen.(1) Goldsmiths, in particular, had strong safes and could offer protection for a merchant's money while he was trading. So, merchants would deposit their money with goldsmiths who would charge them a small fee and issue them with a receipt.(2) These receipts represented a form of debt and, eventually, this debt became 'transferable' in the sense that it became possible for a merchant who wished to make a purchase to transfer the debt to the seller as payment for his goods. Final settlement occurred when the sellers went back to the goldsmith to call in the debt.

---

[1] Indeed, this story also holds true in an economy in which government-issued fiat money is circulating.
[2] One such form of receipt was the 'bill of exchange' on which the goldsmith would promise repayment of the money on a particular date at a particular (distant) location.

Now, one way of making debt transferable was to issue it to 'the bearer'; such debt acted as private 'banknotes'. Of course, private banknotes are just as subject to theft as commodity money or public banknotes. So, an alternative was for the goldsmith to issue the merchant with debt that required the merchant's signature in order for it to be transferred. As emphasised by Kahn and Roberds (2002), since either form of debt was negotiable – that is was issued and transferred according to a clear set of rules concerning the liabilities of buyers, sellers and third-party debtors at all points in the transaction chain – they both had the advantage of offering a means of final payment; in other words, they became 'money' (inside money) themselves.

He *et al* (2005) develop a model of banks based on these ideas. The idea is that merchants deposit their (outside) money with banks who issue them with transferable debt that requires signing over (what they refer to as 'cheques'). What follows is a description of their basic model – 'exogenous theft' – though they discuss a number of extensions in their paper. The environment is much the same as described in the model of Section 2, above, so I will only comment on what is different. The key change is that a proportion, λ, of the agents not holding money are 'thieves'.[3] In a meeting between an agent carrying money and a thief, the thief attempts to steal the buyer's unit of money. He is successful with probability γ and, if he succeeds, incurs a cost z. This cost is motivated by a desire to capture in a simple way the idea that theft imposes negative externalities on society as a whole.[(4)] If there were no cost of theft, then theft would simply represent a transfer of resources and have a zero effect on aggregate welfare. The only other change is that I assume that the probability of a double coincidence of wants, *xy* in Section 2, above, is zero and, so, barter trade never occurs.

The Bellman equations for an agent holding money and one not holding money are, respectively,

$$rV_1 = x(1-M)(1-\lambda)[u+V_0-V_1]+\gamma\lambda(1-M)[V_0-V_1] \quad \textbf{(11)}$$

$$rV_0 = xM(1-\lambda)[V_1-V_0-c]+\gamma\lambda M[V_1-V_0-z] \quad \textbf{(12)}$$

In a monetary equilibrium both types of agent must prefer to be active in the economy rather than drop out and live in autarchy. This implies: $V_1 \geq 0$ and $V_0 \geq 0$. What is more, when a producer meets a buyer who wants his good, the producer needs to have an incentive to accept a unit of money as payment. This implies $V_1 - V_0 \geq c$. Given these conditions, I need only check that $V_0 \geq 0$ and $V_1 - V_0 \geq c$ because when these conditions both hold it must be that $V_1 \geq 0$.

As before, the autarky equilibrium in which no-one accepts money will always exist but, in addition, a monetary equilibrium will exist if the gains from trade are sufficiently large. In particular, He *et al* (2005) prove that monetary equilibria exist if and only if $\frac{x(1-m)(1-\lambda)u+\gamma\lambda zm}{r+x(1-m)(1-\lambda)+\gamma\lambda} \geq c$ and $\frac{(1-m)[x(1-\lambda)+\gamma\lambda]}{r+(1-m)[x(1-\lambda)+\gamma\lambda]}u - \frac{\gamma\lambda}{x(1-\lambda)}z \geq c$.

---

[3] Here, I interpret 'money' as publicly-issued banknotes though it can be easily extended to encompass commodity money and/or privately-issued banknotes.

[4] A part of this cost of theft can also be thought as a proxy for the deadweight cost of using cash. By this, I am thinking of 'cash handling' costs, which can be significant in practice, as well as the actual costs of producing notes and coin.

The next step in building up their model is to add banks. Following He *et al* (2005), I allow agents to deposit their money into banks for safekeeping. The bank issues them with a cheque book and charges a fee, ϕ, for the provision of this service. The key advantage of using cheques, rather than cash, is that they cannot be stolen.[5] In providing the service, they incur a cost, a, per unit of money deposited. I assume that banks are subject to 100% reserve requirements and, so, do not issue loans and always redeem cheques drawn upon them. Perfect competition in the banking sector then implies that $\phi = a$.

Following He *et al* (2005), I assume that each period consists of two sub-periods: 'day' and 'night'. The model as described to date consists of the actions of agents at night. During the day, agents produce and consume a 'general good' that is nonstorable. Utility is linear in the consumption of this good for all agents. The purpose of introducing this general good is to allow, in effect, the transfer of utility; I need to do this in order that agents can pay the fees they owe their banks, which they do in terms of general goods. Since trade in the decentralised market is anonymous, agents cannot use claims on future general goods in the decentralised market unless these are claims drawn on a bank. It is also during the day that agents redeem their cheques, either by opening up an account at the bank of the agent who paid them for their good the previous night, or by walking the money from this bank to their bank. I assume that this can be done with perfect safety (perhaps because the banks are so close together or because the transfer is done with heavy security).

Let θ be the probability that an agent with money deposits it in the bank, V_{d} be the value of an agent with money in the bank, exclusive of the fee (equal to a as said above) and V_{m} be the value of an agent who chooses to carry cash. Then $V_1 = \max\{V_{\mathrm{m}}, V_{\mathrm{d}} - a\}$. The Bellman equations are as follows:

$$rV_d = x(1-M)(1-\lambda)(u+V_0-V_m)+V_1-V_d \quad \textbf{(13)}$$

$$rV_m = x(1-M)(1-\lambda)(u+V_0-V_1)+\gamma\lambda(1-M)(V_0-V_1)+V_1-V_m \quad \textbf{(14)}$$

$$rV_0 = xM(1-\lambda)(V_1-V_0-c)+\gamma\lambda M(1-\theta)(V_1-V_0-z) \quad \textbf{(15)}$$

Now, depending on the values of the parameters, we could identify four possible equilibria:

a) Autarchy: $V_1 - V_0 - c < 0$ and $V_0 < 0$

b) Cash: $\theta = 0$, $V_m \geq V_d - a$, $V_1 - V_0 - c < 0$ and $V_0 \geq 0$

c) Mixture: $0 < \theta < 1$, $V_m = V_d - a$, $V_1 - V_0 - c < 0$ and $V_0 \geq 0$

d) Payment system: $\theta = 1$, $V_m \leq V_d - a$, $V_1 - V_0 - c < 0$ and $V_0 \geq 0$

Note that, as I said earlier, autarky will always be an equilibrium. He *et al* (2005) Proposition 2 shows the conditions under which each of the other equilibria exist. They show that if a

---

[5] To be more precise, the point is not that cheque books cannot be stolen, rather that they cannot be used by anyone except the particular individual on whose account they draw. This means that there is no incentive for thieves to rob any agent who was not carrying cash.

monetary equilibrium exists without banks, it will still exist with banks; moreover, the addition of banks widens the set of parameters over which a monetary equilibrium will exist. In other words, the existence of safe places for storing money will encourage its use in situations where otherwise the presence of thieves would have rendered it too risky to accept. Furthermore, He *et al* find that a lower supply of money makes it more likely that agents will use banks. In this light we can see the development of banks partly as a result of the need to economise on the use of money – the collateral asset.

But a way of economising even more on the use of outside money was to issue even more inside money; that is, for banks to start issuing loans to their customers. The model of Kiyotaki and Moore (2000) shows how the addition of inside money – via the conversion of private debt arrangements into bank debt – can substitute for lack of collateral in a limited commitment economy. In particular, in their model they have farmers who need to borrow in order to make investments that pay off two periods later but are unable to commit all of their future output towards paying off the loan. Furthermore, they assume that debt issued by these agents is non-transferable. These frictions lead to underinvestment, lower than optimal output, a 'jagged' path for consumption – in particular, involving overconsumption at the time the investment projects produce output – and a negative 'liquidity premium' on long-term debt. Kiyotaki and Moore then add 'banks' to the model. In their model, banks issue transferable debt (inside money) against the non-transferable debt issued by the farmers. The presence of banks enables better consumption smoothing, albeit at a direct output cost since they require capital, which depreciates, to operate but produce no output themselves.

In our context, we might think of Kiyotaki and Moore's (2000) 'farmers' as merchants whose 'investments' involve paying their foreign suppliers and whose payoff occurs when they sell their wares to final consumers. Without banks, a merchant would need to issue his own (non-transferable debt) in order to raise the funds to pay his suppliers. Final consumers would then pay with outside money, and the merchant pay off his creditors with this outside money. With banks, the merchant writes cheques to his suppliers. These cheques represent transferable debt of the bank. Allowing the merchant to write cheques implies the granting of an overdraft facility; this can be thought of as a non-transferable debt issued by the merchant. In this way the bank has changed a non-transferable debt into a transferable one. Two periods on, the merchant sells his wares to consumers and deposits the proceeds in his bank. At the same time, the suppliers cash their cheques at the bank, which pays up using the outside money deposited by the merchant.

Kahn and Roberds (2002) have a similar story. Again, a merchant needs to make a payment to a supplier and he can do this by issuing his own (non-transferable) debt. He turns the supplier's goods into final output that he sells on to final consumers. The difference between their model and that of Kiyotaki and Moore (2000) is that, in this case, consumers issue debt to pay the merchant. If this debt is non-transferable, the merchant will have to wait until it is paid off before paying off his supplier; if it is transferable – that is the consumers have used cheques or banknotes to pay the merchant – the merchant pays off the supplier immediately and the supplier eventually calls in the debt from the consumers. The key point in both models is that, where three or more parties form a 'credit chain', the transfer of debt of the most reliable party can affect 'final' settlement. In addition, both models suggest that, by allowing agents to economise

on collateral (including outside money), the introduction of inside money can improve upon allocations involving only outside money.

But, eventually, the circulation of inside money still leads to a requirement for the presence of outside money. In particular, when cheques are cashed or payments are 'walked' from one bank to another bank, outside money still needs to be present. This fact, and the general shortage of outside money, encouraged the banks to develop sophisticated 'banking procedures and … interbank funds transfer systems' (i.e., payment arrangements) in order to economise on the need for it. The development of payment systems is tackled in the next section of the paper.

4 Clearing, settlement and central banking

In describing how interbank funds transfer occurs we need to consider the settlement agent, the settlement institution and the ultimate settlement asset. The BIS defines the settlement agent as an institution that manages the settlement process (e.g., the determination of settlement positions, monitoring of the exchange of payments, etc.) for transfer systems or other arrangements that require settlement. I refer to this process as 'clearing'. The BIS defines the settlement institution as the institution across whose books transfers between participants take place in order to achieve settlement within a settlement system. I refer to this process as 'settlement'.

We could imagine, at least, three models for clearing and settlement:

- A series of bilateral arrangements involving each pair of banks determining their bilateral positions and settlement in outside money

- A mutually-owned clearing house calculating the multilateral net positions of each bank and settlement in outside money. (In practice, this might involve the banks lodging outside money with the clearing house and the clearing house transferring the claims to this outside money, similar to the way gold is moved between different piles held at the Federal Reserve Bank of New York to effect some international payments today.)

- A central bank – one of the banks creates accounts for the other banks and what are transferred are the liabilities of this one bank. In other words, the liabilities of this bank become the ultimate settlement asset – 'outside money'. In this situation, the multilateral net positions of each bank could still be calculated by a clearing house that was operationally and legally separate to the central bank. The key is that final settlement is effected in central bank money. Furthermore, the universal acceptability of this bank's liabilities meant that it could temporarily expand these as a way of easing liquidity crises without their losing value. In other words, this bank was in a position to act as the 'lender of last resort'.[6]

---

[6] Gorton and Huang (2002a) showed that a mutually-owned clearing house could act as a lender of last resort issuing its own liabilities against the mutualised assets of the clearing house members. Indeed, the New York Clearinghouse did just that in the late 19th century in the United States. But Gorton and Huang (2002b) suggest that central banks are necessary since they mitigate the externality of disruption to transactions that is associated with bank panics. Indeed, it was this problem – exemplified by the banking panic of 1907 – that led to the formation of the Federal Reserve system in the United States.

As argued earlier, such arrangements developed as a response to the general shortage of outside money (collateral assets). Now, one immediate way of reducing the money (collateral) needed to effect payments is to use netting. In other words, banks realised that it was cheaper not to settle payments in full as and when they arose but, rather, keep a 'running tab' showing how much they owed each other bank net of what each other bank owed them. Eventually, they moved to a situation where they submitted all the individual payments to a central 'clearer' that could calculate the net amount due to/from each bank from/to all other banks (as opposed to each of the others). Final settlement – walking the money from one bank to another – occurred either after a set period of time, typically the end of each day, or when the net balance became larger than a certain amount (bilateral or multilateral credit limit).

It is straightforward to show that the amount of money (collateral) needed to effect final settlement in a multilateral net settlement system is smaller than the amount needed to effect final settlement in a bilateral net settlement system and this, in turn, is smaller than the amount needed to effect final settlement in a gross settlement system. Kahn, McAndrews and Roberds (2003) examine the advantages of net settlement systems and find that, in addition to reducing the need for collateral, they reduce the incentives for banks to default on their payment obligations where payment occurs with a lag relative to delivery. Hence, a move from gross settlement to net settlement can lead to an increase in trade.

In what follows, I concentrate on the netting benefits. To model these, let there be *n* banks. On day *t*, bank *i* receives notes drawn on bank *j* equal in value to $X_{ji,t}$. So, the net position of bank *i* vis-à-vis bank *j* for day t is $X_{ij} - X_{ji}$. I assume that, although bank *i* knows $X_{ji,t}$ he cannot observe $X_{ij,t}$ until bank *j* presents banks *i*'s notes for final settlement. In order to make the netting benefits clear (and to abstract from issues about the precise timings of payments in a real-time gross settlement system affecting the amount of liquidity needed) I assume that, in the absence of a netting arrangement, bank *i* would need to hold collateral equal to $\sum_j X_{ij,t}$ in order to make all his payments on day *t*. The expected benefit to bank *i* of agreeing to a bilateral netting arrangement with bank *j* would then be $Min\left(E\left(X_{ij}\right), E\left(X_{ji}\right)\right)$. Following this argument through, bank *i* would gain from bilateral netting arrangements with all banks whose notes it was receiving.

Now suppose that all *n* banks got together to form a multilateral arrangement. In particular, suppose that the banks formed a mutually-owned clearing house that carried out the netting and effected final settlement in outside money. The netting benefits obtained by bank *i* within such a system would be $Min\left(E\left(\sum_j X_{ij}\right), E\left(\sum_j X_{ji}\right)\right) \geq \sum_j Min\left(E\left(X_{ij}\right), E\left(X_{ji}\right)\right)$ where the inequality is strict provided that bank *i* is a net payer of at least one other bank and a net payee of at least one other bank. Aggregating this up implies that the multilateral system will require less collateral than a set of bilateral arrangements provided that at least two banks are either both net payers or net payees. In addition, a multilateral arrangement is likely to be less costly than a set of bilateral

arrangements since it only involves only *n* transfers of outside money as opposed to $\frac{n(n-1)}{2}$ transfers.

The next question to answer is how the development of clearing houses are related to the development of central banks. In the situation discussed above, the clearing house acts solely to calculate the net amounts due from/to each bank. Final settlement still involves the movement of collateral (outside money), albeit across the floor of the clearing house's vaults. Of course, this in itself represented an efficiency gain as collateral stored in the clearing house vaults would be made secure from theft than collateral being transported between banks. The key difference between this system and one involving a central bank is that final settlement is enacted via the transfer of balances across the books of this bank; that is, the liabilities of this bank become the settlement asset. And the key advantage of this is that the central bank need not hold any of the original collateral asset to back its liabilities (though, of course, there would be an issue about whether the other banks would be happy to use its liabilities to settle payments if they were not backed by 'safe' assets).

Again I can calculate the collateral savings. For and bank *i* that is not the central bank, the savings will still be $Min\left(E\left(\sum_j X_{ij}\right), E\left(\sum_j X_{ji}\right)\right)$. For the central bank (bank *k*, say), the savings will be $E\left(\sum_j X_{kj}\right) \geq Min\left(E\left(\sum_j X_{kj}\right), E\left(\sum_j X_{jk}\right)\right)$. In other words, an arrangement involving a central bank will achieve at least as large a netting benefit as in a multilateral arrangement with settlement in outside money and a greater benefit when the central bank is a net payer. The number of transfers required is the same in each of the two systems suggesting that a central bank arrangement is no more costly than a multilateral clearing house arrangement.

In terms of the economics of social networks, I have shown that a payments network involving multilateral clearing and a central bank is 'efficient' in the sense that the total benefits accruing to the n banks is larger than under any alternative network. This leaves the question of whether such a network is stable. Following Jackson (2005), I define a 'pairwise stable' multilateral payments arrangement to be one in which no bank wishes to leave the arrangement and no two banks wish to deal bilaterally with each other.

To check whether the central bank arrangement is pairwise stable, consider non-central banks *i* and *j* that are members of the system. Suppose they make no payments to and from each other. Then setting up a bilateral arrangement would incur some cost and yield no benefit. If they do make payments to and from each other, creating a bilateral link and shifting their payments out of the multilateral arrangement will never lead to a reduction in their collateral needs, while still imposing some cost. The central bank will be linked to every bank it deals with and so setting up a bilateral arrangement with any other bank would incur some cost and yield no benefit. So, this network is stable against the addition of links. Now consider banks *i* and *k* where bank *k* is the central bank. Delinking will create a need for bank *i* to create an arrangement with another bank *j* for making its payments to all the banks in the network (a 'correspondent banking arrangement').

In this case, bank *i* will need the same amount of collateral as in the original multilateral arrangement but bank j will need collateral equal to $Max\left(0, E\left(\sum_l X_{il} - X_{li}\right) + E\left(\sum_l X_{jl} - X_{lj}\right)\right)$ $\geq Max\left(0, E\left(\sum_l X_{jl} - X_{lj}\right)\right)$. The total number of links will be the same suggesting a similar cost between the two arrangements. So, this network is stable against the removal of links.

The problem with the network I have described above is that it is, in general, not unique. Any bank that is a net payer is a candidate for being the central bank. One way of narrowing the set of possible equilibria is to impose the entire cost of supporting the network onto the central bank; in the multilateral netting system involving a clearing house the cost would likely be born jointly across all the banks. Suppose this cost is *c*. Then a bank (bank *k*, say) would only wish to become the central bank if $E\left(\sum_j X_{kj}\right) - Min\left(E\left(\sum_j X_{kj}\right), E\left(\sum_j X_{jk}\right)\right) \geq \frac{c(n-1)}{n}$.

This all suggests a need to adapt the model so as to explain why one particular bank, rather than another, might become the central bank and why such a bank would wish to become the central bank given that it is costly to do so.

One answer to the question suggested by the analysis above is that a bank that was a large net payer is most likely to become the central bank. But, if this were the case, the bank's liabilities would be increasing over time. The equilibrium would only be sustainable if the other banks remained happy to use the central banks liabilities as the settlement asset and this, in turn, implies a need for these liabilities to not grow to fast (that is inflation to be low) and to be backed by 'safe' assets that could easily be grown over time. Unlike other banks in the United Kingdom at the time, the Bank of England – as a joint stock company – was able to raise capital without having to rely on just a handful of individuals. This enabled it to grow its assets. In addition, the bulk of its assets were government bonds – seen as particularly safe. Finally, since it was the government's bank, it was, at least initially as a result of the Napoleonic Wars, a large net payer. In the context of the model, these factors would suggest that the Bank of England would emerge as the central bank. For such central banks, the granting of 'legal tender' to its money (as happened to Bank of England money in 1833, for example) merely confirms the position of ultimate settlement asset that it has already achieved.

But more often there was no bank that could fit these criteria and emerge as the central bank. In these cases, the alternative involved the government simply creating a central bank and paying for the cost of running the system out of general taxation. In continental Europe, 'Public Banks' (as they were called) grew up as a regulatory answer to the problem of large-scale bank failure. (See Kohn, (1999) and Quinn and Roberds (2005).) In particular, various governments felt the need to set up institutions specialising in payments, that is, not engaging in intermediation and so not subject to failure. As these institutions were not able to make a profit, their costs were paid for by government. Examples of public banks that were successful include the Banco di Rialto in Venice and the Bank of Amsterdam (after which the Bank of England was modelled). More recently, governments have set up central banks at the same time as which they have created a fiat money – the money of their central bank – and enforced legal tender properties on it. Given

this, the central bank money became the obvious settlement asset in these countries. There are many examples of this including the Federal Reserve System in the United States, the Bank of Canada and, most recently, the European Central Bank.

As discussed in the motivation for this paper, central banks around the world today generally share two core purposes – the provision of monetary and financial stability. But these responsibilities arose naturally once the bank's liabilities became the ultimate settlement asset.

To see this, I first note that once this happened, this bank could elastically increase the supply of its liabilities in times of crisis to enable payments to happen; in other words, it becomes able to act as the lender of last resort.(7) In addition, it had an incentive to do this since if it allowed a solvent commercial bank to fail as a result of a run, it would only aggravate the situation and this could ultimately result in a run on itself. Put differently, profit maximisation is consistent with Bagehot's (1873) rule that a central bank should always lend to liquid but solvent institutions against collateral. This also creates an incentive for the central bank to minimise the likelihood of this happening, that is, to ensure a generally stable financial system (financial stability). Furthermore, this will create a need for the central bank to monitor individual banks so as to ensure that it does not lend to insolvent banks. However, if it is taxpayers' money that is used to finance such lending, the incentive is for the government to ensure that banks are properly monitored; whether this is done by the central bank or a different authority (such as the Financial Services Authority in the United Kingdom) is a choice for the government.

In addition, as discussed above, privately-owned providers of the ultimate settlement asset also have incentives to maintain monetary stability by ensuring that their liabilities do not grow at too fast a rate. In particular, if it printed more and more of its notes without a corresponding increase in the demand for them, the notes would fall in value relative to those of other banks. Eventually, it would no longer be seen as 'safe' and it would lose the revenue it obtained from acting as the settlement institution. One way of ensuring that its notes were seen as 'safe' was to obtain a government guarantee – which could take the form of its notes being declared as legal tender. In that case, it would be able to increase the circulation of its notes without affecting their acceptability; in turn, this would enable the central bank to ensure that payments were made even in adverse circumstances. In this case, the bank would be likely to try and make profits from seignorage, provided the government allowed it to keep these profits and did not appropriate them for itself.(8) Profit maximisation would likely lead to a rate of inflation that was low and stable, though it is also likely that this would be higher than optimal.

---

[7] The Federal Reserve was set up to do just this. The full title of the Federal Reserve Act of 1913 reads, 'An act to provide for the establishment of Federal reserve banks, *to furnish an elastic currency*, to afford means of rediscounting commercial paper, to establish a more effective supervision of banking in the United States, and for other purposes' (my italics).

[8] For instance, as a result of the Bank of England Act of 1844, the UK government is entitled to the seignorage revenue made from the production of Bank of England notes.

5 Conclusions

In this review essay, I have attempted to explain the economics behind the evolution of payments. In so doing, I first assessed why the introduction of money – by which I mean either commodity money or pure fiat money issued by a government – into a barter economy can be welfare improving. Following, *inter alia*, Kiyotaki and Wright (1993) and Kocherlakota (1998) I argued that money can overcome the 'double coincidence of wants' problem in an economy in which agents are unable to commit to repaying loans and their trading histories are private information. Following, *inter alia*, He, Huang and Wright (2005) and Kohn (1999) I then argued that the introduction of banks could further improve the situation by addressing the twin problems that it is hard to assess the quality and legitimacy of money and that money is subject to theft.

But even with banks there would still be a need for some money to be present to effect final settlement. The general shortage of money in the middle ages encouraged the banks to develop sophisticated 'banking procedures and … interbank funds transfer systems' (i.e., payment arrangements) in order to economise on the need for it. Payment systems typically reduced the need for money via netting: initially bilateral netting between pairs of banks but eventually multilateral netting across a group of banks. The netting would be carried out by a mutually-owned clearing house and final settlement would be effected via the transfer of outside money.

Having described the benefits to such systems, I then examined the further improvement that could be offered if the settlement was carried out across the books of one of the banks: a 'central bank'. I argued that, in addition to a further saving of collateral, once this happened it became possible for the central bank to act as a lender of last resort and that, in turn, this led it to become concerned about financial and monetary stability. In other words, I argued that central banks developed to perform a payments role and that their core purposes of monetary and financial stability stemmed from their performing this role.

So, what are the core roles of a central bank in payment systems? Clearly, where banks require final settlement in an asset that they deem to be 'as good as gold', the liabilities of the central bank should be the settlement asset. Where there liabilities are being used, it is clear that they need to exercise control over them; banks would soon stop settling in the liabilities of the central bank if the central bank were seen to have no control over them. But there is nothing in the analysis that suggests the central bank should either own or operate the systems (that is, carry out the 'clearing') subject to maintaining control over the supply of its liabilities. Having said that, the lender of last resort function suggests that in times of general liquidity shortages, the central bank needs to be able to get its liabilities out to those banks that need them. Perhaps the most efficient way of ensuring this is for the central bank to be a member of any system where their liabilities were used as the settlement asset. But it also suggests that the central bank needs to ensure that these systems keep operating in times of liquidity shortages, either by operating the system itself or by having the right to step in and operate the system at such times.

References

Bagehot, W (1873), *Lombard Street: A description of the money market*, London: Henry S. King.

Baumol, W J, and Blinder, A S (1991), *Macroeconomics: Principles and policy*, (5th Edition), London: Harcourt Brace Jovanovich.

Gorton, G and Huang, L (2002a), 'Bank panics and the endogeneity of central banking', National Bureau of Economic Research *Working Paper* No. 9102.

Gorton, G and Huang, L (2002b), 'Banking panics and the origin of central banking', National Bureau of Economic Research *Working Paper* No. 9137.

Grierson, P (1977), *The origins of money*, London.

He, P, Huang, L, and Wright, R (2005), 'Money and banking in search equilibrium', *International Economic Review*, Vol. 46, pages 637-70.

Jackson, M O (2005), 'The economics of social networks', California Institute of Technology, *mimeo*.

Kahn, C, McAndrews, J, and Roberds, W (2003), 'Settlement risk under gross and net settlement', *Journal of Money, Credit, and Banking*, Vol. 35, No. 4, pages 591-608.

Kahn, C, and Roberds, W (2002), 'The Economics of Payment Finality', Federal Reserve Bank of Atlanta *Economic Review*, Vol. 87, pages 30-39.

Kiyotaki and Moore (2000), 'Inside money and liquidity', London School of Economics, *mimeo*.

Kiyotaki, N, and Wright, R (1993), 'A search-theoretic approach to monetary economics', *American Economic Review*, Vol. 83, No. 1, pages 63-77.

Kocherlakota, N (1998), 'Money is memory', *Journal of Economic Theory*, Vol. 81, pages 232-51.

Kohn, M (1999), 'Early deposit banking', Dartmouth University *Working Paper* No. 99-03.

Quinn, S, and Roberds, W (2005), 'The big problem of large bills: The Bank of Amsterdam and the origins of central banking', Federal Reserve Bank of Atlanta *Working Paper* No. 2005-16.

Rosenblat, T S (1999), 'What makes the money go round?', Massachusetts Institute of Technology *mimeo*.

Rupert, P, Schindler, M, Shevchenko, A, and Wright, R (2000), 'The search-theoretic approach to monetary economics: A primer', University of Pennsylvania *mimeo*.

Velde, F, Webber, W, and Wright, R (1999), 'A model of commodity money with applications to Gresham's Law and the debasement puzzle', *Review of Economics Dynamics*, Vol. 2, pages 291-323.