

THE EVOLUTION OF A FREE BANKING SYSTEM*

with Lawrence H. White

INTRODUCTION

Monetary theorists' assumptions concerning the institutional features of a completely unregulated monetary system have ranged from the proliferation of numerous competing private fiat currencies at one extreme to the complete disappearance of money at the other.¹ While these assumptions have generated clear-cut and provocative conclusions, their plausibility or realism in light of historical experience is open to serious doubt. These doubts may unfortunately suggest that any discussion of an unregulated monetary system (or "free banking" system) must be tenuous and highly speculative. This chapter tries to show, to the contrary, that important institutional features of a free banking system, in particular the nature of payment media, can be realistically grounded by constructing a logical explanation of its evolution.

The method of logical evolutionary explanation has previously been applied to monetary institutions by Hicks (1967) and Menger (1892), among others. The present study integrates and extends work along their lines. The method is employed here in the belief that it has been unduly neglected in recent work, not that it is the only valid method for theoretically explaining institutional arrangements. The more standard method of building explicit transactions costs or informational imperfections or asymmetries into an optimization model has unquestionably been useful in the task of explaining why banks exist as intermediaries (Santomero (1984, 577-80) surveys this literature).

Our investigation derives arrangements that would have arisen had state intervention never occurred. The results should therefore help to identify the degree to which features of current monetary and banking institutions are rooted in market forces and the degree to which they have grown out of regulatory intervention. Such information gives important clues about how

*Reprinted, with permission, from *Economic Inquiry*, Vol. XXV, No. 3 (July 1987).

future deregulation would modify institutions. We show that sophisticated monetary arrangements, whose institutional features are described, emerge in the absence of regulation. No strong claims are advanced here about the welfare properties of these arrangements.² We aim to establish the most credible path for unrestricted monetary evolution, but certainly not the only possible path. Economists who find other institutional outcomes more plausible for an unregulated system will, we hope, similarly try to explain why and how those outcomes would emerge.

The evolution of a free banking system, following the emergence of standardized commodity money, proceeds through three stages. These are, first, the development of basic money-transfer services which substitute for the physical transportation of specie; second, the emergence of easily assignable and negotiable bank demand liabilities (inside money); and third, the development of arrangements for the routine exchange ("clearing") of inside monies among rival banks. The historical time separating these stages is not crucial. The path of development, rather than being one of steady progress as pictured here, may in practice involve false starts or creative leaps. What is essential is that each stage is the logical invisible-hand outgrowth of the circumstances that preceded it. In other words, each successive step in the process of evolution originates in individuals' discovery of new ways to promote their self-interest, with the outcome an arrangement at which no individual consciously aims.

COMMODITY MONEY

Because the use of money logically and historically precedes the emergence of banking firms, we begin with an account of the origin of money. Our account follows that of Menger (1892), who furnished an invisible-hand explanation, consistent with historical and anthropological evidence, of how money originated as a product of undesigned or spontaneous evolution.³ Menger's theory shows that no state intervention is necessary in order to establish a basic medium of exchange or unit of account. It also provides a useful prototype for our explanations of how subsequent banking institutions evolve in spontaneous fashion.

In premonetary society, traders relying upon barter initially offer goods in exchange only for other goods directly entering their consumption or household production plans. The number of bargains struck this way is small, owing to the well-known problem of finding what Jevons termed a "double coincidence of wants." Before long some frustrated barterer realizes that he or she can increase his or her chances for success by adopting a two-stage procedure. The barterer can trade his or her wares for some good, regardless of its direct usefulness which will more easily find a taker among those selling what he or she ultimately wants. It follows that the earliest media of exchange are simply goods perceived to be in relatively

widespread demand. The widening of demand for these things owing to their use as media of exchange reinforces their superior salability. Other traders eventually recognize the gains achieved by those using indirect exchange, and emulate them, even though they may be unaware of the reason for the advantages from using a medium of exchange. This emulation further enhances the acceptance of the most widely accepted media, elevating one or two goods above all others in salability. The snowballing of salability results in the spontaneous appearance of generally accepted media of exchange. Eventually traders throughout an economy converge on using a single commodity as a generally accepted medium of exchange, i.e., as money.

Historical evidence on primitive monies indicates that cattle were often the most frequently exchanged commodity, and that a standardized "cow" was the earliest unit of account. Cattle were a poor general medium of exchange, however, because of their relative nontransportability and non-uniformity. Not until the discovery of metals and of methods for working them did the use of money replace barter widely.⁴ According to Jacques Melitz (1974, 95), common attributions of moneyness to primitive media, especially nonmetallic "moneys" (with the exception of cowries in China), warrant skepticism because many of these media (e.g., the Yap stones of Melanesia) do not meet any reasonably strict definition of money.

The emergence of coinage can also be explained as a spontaneous development, an unplanned result of merchants' attempts to minimize the necessity for assessing and weighing amounts of commodity money received in exchange. Merchants may at first mark irregular metallic nuggets or pieces after having assessed their quality. A merchant recognizing his or her own or another's mark can then avoid the trouble and cost of reassessment. Marking gives way to stamping or punching, which eventually leads to specialists' making coins in their modern form. Techniques for milling coin edges and covering the entire surface with type provide safeguards against clipping and sweating and so allow coinage to serve as a guarantee of weight as well as of quality. Arthur R. Burns (1927a, 297–304; 1927b, 59) has illustrated this process with evidence from ancient Lydia, where coins of electrum (a naturally occurring silver–gold alloy) came into early use.

Absent state interference, coinage is a private industry encompassing various competing brands. Under competition coins are valued according to bullion content plus a premium equal to the marginal cost of mintage. The demand for readily exchangeable coins promotes the emergence of standard weights and fineness. Nonstandard coins must circulate at a discount because of the extra computational burden they impose, so that their production is unprofitable. States seem to have monopolized coinage early in history, but not by outcompeting private mints. Rather, the evidence suggests that state coinage monopolies were regularly

established by legal compulsion and for reasons of propaganda and monopoly profit. State-minted coins functioned both as a symbol of rule and as a source of profits from shaving, clipping, and seignorage. For these reasons coinage became a state function throughout the world by the end of the seventh century (Burns 1927a, 308; 1927b, 78).

BANKING FIRMS

The counting and transporting of coins entail considerable inconvenience. Traders, particularly those frequently making large or distant exchanges, will naturally seek lower-cost means of transferring ownership of money. One likely locus for development of such means is the market where local coins are exchanged for foreign coins. Standard coins may differ interlocally even in the absence of local state interventions because of geographic diseconomies in reputation building for mints. A coin-exchange market then naturally arises with interlocal trade. A trader who uses a money changer must initially count and carry in local coin each time he or she wants to acquire foreign coin, or vice versa. The trader can reduce costs by establishing a standing account balance, to build up at his or her convenience and draw upon as desired. The money changer's inventories equip him or her to provide such accounts, which constitute demand deposits, and even to allow overdrafts. These deposits may originally be nontransferable. But it will soon be apparent, where one customer withdraws coins in order to pay a recipient who redeposits them with the same exchange banker, that the transfer is more easily made at the banker's place of business, or more easily yet by persuading the banker to make the transfer on his or her books without any handling of coins. Thus trading individuals come to keep money balances with agencies which can make payments by ledger-account transfers.

Money-transfer services of this sort, provided by money changers and bill brokers in twelfth century Genoa and at medieval trade fairs in Champagne, mark the earliest recorded forms of banking.⁵ In time all the major European trading centers had "transfer banks," as Raymond de Roover (1974, 184) calls them; he comments that "deposit banking grew out of [money-changing] activity, because the money changers developed a system of local payments by book transfer." In our view, however, the taking of deposits on at least a small scale logically *precedes* the development of book-transfer methods of payment.

Money-transfer services may also develop in connection with deposits made for safekeeping rather than for money changing. The well-known story of the origins of goldsmith banking in seventeenth century England illustrates this development. Wealthy persons may temporarily lodge commodity money with scrivemers, goldsmiths, mintmasters, and other reputable vault-owners for safekeeping. Coin and bullion thus lodged

must be physically withdrawn and transferred for its owner to use it as a means of payment. Exchanges in which the recipient redeposits it in the same vault (like redeposits with a money changer or bill broker) create obvious advantages in making the transfer at the vault, or better yet in simply notifying the vault's custodian to make the transfer on his or her books. In England, scriveners were the earliest pioneers in the banking trade; in Stuart times they were almost entirely displaced by goldsmith bankers. English goldsmiths evidently became transfer bankers during the seventeenth century, when they "began to keep a 'running cash' for the convenience of merchants and country gentlemen" (de Roover 1974, 83-4). The confiscation by Charles I of gold deposited for safekeeping at the Royal Mint ended that institution's participation in the process of banking development. Private mints, had they been permitted, would have been logical sites for early banking activities.

Transfer banking is not connected with intermediation between borrowers and lenders when the banker acts strictly like a warehouse, giving deposit receipts which are regular warehouse dockets. The strict warehouse banker is a bailee rather than a debtor to depositors and can make loans only out of his or her personal wealth. Two conditions make it possible, however, to take advantage of the interest income available from lending out depositors' balances, even while satisfying depositors' desire to have their funds withdrawable on demand: (1) money is fungible, which allows a depositor to be repaid in coin and bullion not identical to that brought in, and (2) the law of large numbers with random withdrawals and deposits makes a fractional reserve sufficient to meet actual withdrawal demands with high probability even though any single account may be removed without notice. (Interestingly, these conditions may also be met in the warehousing of standard-quality grain, so that fractional reserve "banking" can likewise develop there, as Williams (1984) has shown.) The lending of depositors' balances is an innovation that taps a vast new source of loanable funds and alters fundamentally the relationship of the banker to the depositor customers.

Historically in England, according to Richards (1965, 223), "the bailee . . . developed into the debtor of the depositor; and the depositor became an investor who loaned his money . . . for a consideration." Money "warehouse receipts" became merely ready promissory notes. W. R. Bisschop (1910, 50n) reports that English warehouse bankers had become intermediaries by the time of Charles II (1660-85): "Any deposit made in any other shape than ornament was looked upon by them as a free loan." Competition for deposits prompted the payment of interest on deposits, and the attractiveness of interest on safe and accessible deposits in turn apparently made the practice of depositing widespread among all ranks of people (Powell 1966, 56-7).

THE NATURE OF FREE BANKING

TRANSFERABLE INSTRUMENTS

Under these circumstances the effective money supply obviously becomes greater than the existing stock of specie alone. The most important banking procedures and devices, however, have yet to develop. Many purchases are still made with actual coin. Bank depositors, in order to satisfy changing needs for money at hand, make frequent withdrawals from and deposits into their bank balances. These actions may in the aggregate largely cancel out through the law of large numbers. But they require the banks to hold greater precautionary commodity money reserves, and consequently to maintain a larger spread between deposit and loan rates of interest, than is necessary when payments practices become more sophisticated. Greater sophistication comes with the emergence of negotiable bank instruments, able to pass easily in exchange from one person to another, which replace coin and nonnegotiable deposit receipts in transactions balances. The use of coin is also superceded by the development of more efficient means for the bank-mediated transfer of deposits.

Assignability and negotiability may develop through several steps. Initially the assignment of deposited money (whether “warehoused” or entrusted to the banker for lending at interest) by the depositor to another party may require the presence of all three parties to the exchange or their attorneys. Money “warehouse receipts” (or promissory notes) and running deposit balances cannot be assigned by the owner’s endorsement without the banker acting as witness. An important innovation is the development of bank-issued promissory notes transferable by endorsement. Assignable notes in turn give way to fully negotiable banknotes assigned to no one in particular but instead payable to the bearer on demand. A parallel development is the nonnegotiable check enabling the depositor to transfer balances to a specific party, in turn giving way to the negotiable check which can be repeatedly endorsed or made out “to cash.”⁶ Thus the modern forms of inside money – redeemable bearer banknotes and checkable deposits – are established. Once this stage is reached it is not difficult for bankers to conceive what Hartley Withers (1920, 24) has called “the epoch-making notion” – in our view it is only an incremental step – of giving inside money not only to depositors of metal but also to borrowers of money. The use of inside money enhances both customer and bank profits, so that only the possible reluctance of courts to enforce obligations represented by assigned or bearer paper stands in the way of its rapid development.

In England bearer notes were first recognized during the reign of Charles II, about the time when warehouse banking was giving way to fractional reserve transfer banking. At first the courts gave their grudging approval to the growing practice of repeated endorsement of promissory notes. Then

Parliament. In France, Holland, and Italy during the sixteenth century merchants' checks "drawn in blank" circulated within limited circles and may have cleared the way for the appearance of banknotes (Usher 1943, 189; Richards 1965, 46, 225).

REGULAR NOTE EXCHANGE

Further economies in the use of commodity money require more complete circulation of inside money in place of commodity money, and more complete development of banknote and check clearing facilities to reduce the need for commodity money reserves. It is relatively straightforward to show that bankers and other agents pursuing their self-interest are indeed led to improve the acceptability of inside money and the efficiency of banking operations.

At this stage, although banknotes are less cumbersome than coin, and checkable deposits are both convenient for certain transactions and interest paying, some coin still remains in circulation. Consumers trust a local bank's notes more than a distant bank's notes because they are more aware of the local notes' likelihood of being honored and more familiar with their appearance (hence less prone to accepting forgeries). It follows that the cost to a bank of building a reputation for its issues – particularly regarding note convertibility – is higher in places further from the place of issue and redemption. The establishment of a network of bank branches for redemption is limited by transportation and communication costs. In the early stages of banking development the par circulation of every bank's notes and checks is therefore geographically relatively limited.⁷ People who generally hold the inside money of a local bank but who do business in distant towns must either take the trouble to redeem some of their holdings for gold and incur the inconvenience of transporting coin, or suffer a loss in value on their notes by carrying them to a locale where they are accepted only at a discount, if at all. (The alternative practice of keeping on hand notes from each locality they deal with is likely to be prohibitively costly in terms of foregone interest.) In general, a brand of inside money will initially be used only for transactions in the vicinity of the issuer, and coin will continue to be held alongside notes of like denomination. The use of commodity money in circulation requires banks to hold commodity reserves greater than those required by the transfer of inside money, because the withdrawal of commodity money for spending generates more volatile reserve outflows than the spending of notes or deposits.

In this situation, profit opportunities arise which prompt actions leading to more general acceptance of particular inside monies. The discounting of notes outside the neighborhood of the issuing bank's office creates an arbitrage opportunity when the par value of notes (i.e. their face redemption value in commodity money) exceeds the price at which they can be

THE NATURE OF FREE BANKING

purchased for commodity money or local issues in a distant town plus (secularly falling) transaction and transportation costs. As interlocal trade grows, "note brokers" with specialized knowledge of distant banks can make a business, just as retail foreign currency brokers do today, of buying discounted nonlocal notes and transporting them to their par circulation areas or reselling them to travelers bound for those areas. Competition eventually reduces note discounts to the level of transaction and transportation costs plus a factor for redemption risk. By accepting the notes of unfamiliar banks at minimal commission rates, brokers unintentionally increase the general acceptability of notes, and promote their use in place of commodity money.

To this point we have implicitly assumed that banks refuse to accept one another's notes. This is not unreasonable; banks have as many reasons as other individuals do to refuse notes unfamiliar to them or difficult to redeem. They have in addition a further incentive for refusing to accept notes from rival banks, which is that by doing so they help to limit the acceptability of these notes, thereby enhancing the demand for their own issues. To cite just one historical illustration of this, the Bank of Scotland and the Royal Bank of Scotland – the first two banks of issue located in Edinburgh – refused to accept the notes of "provincial" banks of issue for a number of years (see Checkland 1975, 126).

Nevertheless note brokerage presents opportunities for profit to bankers. Banks can outcompete other brokers because, unlike other brokers, they can issue their own notes (or deposit balances) to purchase "foreign" notes and need not hold costly till money. Each bank has an additional incentive to accept rival notes: larger interest earnings. If the notes acquired are redeemed sooner than the notes issued, interest-earning assets can be purchased and held in the interim. This profit from "float" can be continually renewed. In other words, a bank can maintain a permanently larger circulation of its own notes by continually replacing other notes with its own, and correspondingly can hold more earning assets than it otherwise could. If other banks are simultaneously replacing Bank A's notes with their own, there may be no absolute increase in A's circulation compared to the situation in which no bank accepts rival notes. But there will be an increase compared to Bank A not accepting, given whatever policies rivals are following, so that the incentive remains. (We argue below that in fact an indirect consequence of *other* banks' par acceptance of Bank A notes will be an absolute increase in A-note holding in place of specie holding.) Where transaction and transportation costs and risks are low enough, competition for circulation will narrow the brokerage fee to zero, i.e., will lead the banks to general acceptance of one another's notes at par. The development of par acceptance by this route does not require that the banks explicitly and mutually agree to such a policy.

An alternative scenario, which assumes strategic behavior by the banks, leads to the same result. A bank may aggressively purchase foreign notes in the market, and then suddenly return large quantities to their issuers for redemption in commodity money, hoping to force an unprepared issuer to suspend payments. The aggressor hopes to gain market share by damaging a rival's reputation or even forcing it into liquidation. These tactics, historically known as "note picking" and "note duelling," initially provoke the other issuers to respond in kind. Collecting and redeeming the first bank's notes not only returns the damage, but helps replenish the other banks' reserves. Purchasing its rivals' notes at par allows a bank to collect them in greater quantities, and may therefore be adopted. (Arbitrage-redemption of notes paid out precludes paying a price above par.) In the long run, nonaggression among banks should emerge, being less costly for all sides. Note picking and note duelling are costly and ineffectual ways to promote circulation when others do likewise. Banks thus find it profitable to take rivals' notes only as these are brought to them for deposit or exchange, and to return the collected notes to their issuers promptly in exchange for commodity money reserves. This result is contrary to Eugene Fama's (1983, 19) suggestion that note duelling will persist indefinitely. It is an example of the "tit for tat" strategy, as discussed by Robert Axelrod (1984), proving dominant in a repeated-game setting.⁸ Again, no explicitly negotiated pact is necessary. It only takes a single bank acting without cooperation from other banks to nudge the rest towards par acceptance (zero brokerage fees) as a defensive measure to maintain their reserves and circulation.

In New England at the beginning of the nineteenth century the Boston banks gave the nudge that put the whole region – with its multitude of "country" banks of issue far removed from the city – on a par-acceptance basis (Trivoli 1979). In Scotland the Royal Bank, when it opened for business in 1727, immediately began accepting at par the notes of the Bank of Scotland, at that time its only rival, and instigated a short-lived note duel. One response by the Bank of Scotland, later widely adopted, is notable: the Bank inserted a clause into its notes giving it the option (which it did not normally exercise) of delaying redemption for six months, in which event it would pay a bonus amounting to 5 per cent per annum (Checkland 1975, 60, 67–8). In both places established banks, even after they had begun accepting each other's notes at par, sometimes refused to take the notes of new entrants. They soon changed their policies because the new banks that accepted and redeemed their notes were draining their reserves, while the established banks could not offset this without engaging in the same practice.

Banks that accept other banks' notes at par improve the market for their own notes and, unintentionally, for the notes that they accept. This makes a third scenario possible: if two banks both understand these circulation

gains, they may explicitly enter a mutual par-acceptance arrangement. Others will emulate them, leading to general par acceptance. This explanation, previously offered by White (1984a, 19–21), assumes slightly more knowledge on the part of banks than the first two scenarios. Historical evidence of such explicit arrangements in Scotland is provided by Munn (1975).

Statistics from Boston dramatically illustrate the mutual circulation gains from par-acceptance arrangements. From 1824 to 1833 the note circulation of the Boston banks increased 57 per cent, but the Boston circulation of country banks increased 148 per cent, despite the Boston banks' intent to drive the country banks out of business (Lake 1947, 186; Trivoli 1979, 10–12). There is room for all banks to gain because the spread of par acceptance makes inside money more attractive to hold relative to commodity money. Since notes from one town are now accepted in a distant town at par, there is no longer good reason to lug around commodity money. As par note acceptance developed in Scotland, Canada, and New England – places where note issue was least restricted – during the nineteenth century, gold virtually disappeared from circulation. (Small amounts of gold coin were still used in these places at least in part because of restrictions upon the issue of “token” coin and of small denomination notes. In an entirely free system, such restrictions would not exist.) In England and the rest of the United States, where banking (and note issue in particular) were less free, gold remained in circulation.

Even the complete displacement of commodity money in circulation by inside money does not, however, exhaust the possibilities for economizing on commodity money. Much of the specie formerly used in circulation to settle exchanges outside the banks may still be needed to settle clearings among them. Banks can substantially reduce their prudentially required holdings of commodity money by making regular note exchanges which allow them to offset their mutual obligations. Only net clearings rather than gross clearings are then settled in commodity money. The probability of any given-sized reserve loss in a given period is accordingly reduced (by the law of large numbers) and each bank can prudently reduce its ratio of reserves to demand liabilities.

The gains to be had from rationalization of note exchange are illustrated by the provincial Scottish banks before 1771, which practiced par acceptance without regular exchange. Note duelling among these banks was not uncommon (Leslie 1950, 8–9; Munn 1981, 23–4), and to guard against redemption raids they had to keep substantial reserves. Munn's figures (1981, 141) show that their reserves during this period were typically above 10 per cent of total liabilities. This contrasts with reserve ratios of around 2 per cent that were typical after note clearings became routine. The advantages of regular note exchange are great enough to have secured its

eventual adoption in every historical instance of relatively free plural note issue.

CLEARINGHOUSES

The most readily made arrangements for note exchange are bilateral. In a system of more than two issuers, however, multilateral note exchange provides even greater economies. Reserve-holding economies result from the offsetting of claims that would otherwise be settled in specie. Multilateral clearing also allows savings in time and transportation costs by allowing all debts to be settled in one place and during one meeting rather than in numerous scattered meetings.

The institutional embodiment of multilateral note and deposit exchange, the clearinghouse, may evolve gradually from simpler note-exchange arrangements. For example, the note-exchange agents of banks A and B may accidentally meet each other at the counter of bank C. The greater the number of banks exchanging bilaterally, the less likely it is that such an encounter could be avoided. It would be natural for these two agents to recognize the savings in simple time and shoe-leather costs from settling their own exchange then and there, and from agreeing to do it again next time out, and then regularly. From a set of three pairwise settlements around one table it is not a large step toward the computation and settlement of combined net clearing balances. Once the advantages of this become clear to management, particularly the reserve-holding economies which may not have concerned the note porters, the institution will spread. Fourth, fifth, and subsequent banks may join later meetings. Or similar, regular few-sided exchanges may be formed among other groups of banks, either independently or by one of the first three banks, whose meetings are later combined with the meetings of the original group. Eventually all the banks within an economy will be connected through one or a small number of clearinghouses.

The histories of the best-known early clearinghouses, in London, Edinburgh, and New York all conform to this general pattern. Gibbons (1858, 292) reports that in New York the impetus for change from numerous bilateral exchanges to combined multilateral exchange came from note porters who "crossed and re-crossed each other's footsteps constantly." Among the London check porters, as related by Bisschop (1910, 160), "occasional encounters developed into daily meetings at a certain fixed place. At length the bankers themselves resolved to organize these meetings on a regular basis in a room specially reserved for this purpose."

The settlement of interbank obligations is initially made by physical transfer of commodity money at the conclusion of clearing sessions. Banks will soon find it economical to settle instead by means of transferable reserve accounts kept on the books of the clearinghouse, echoing the

original development of banks or equity shares denominated in currency units. As a transfer bank, the clearinghouse need not hold 100 per cent reserves, and can safely pay its members a return (net of operating costs) by holding safe earning assets. This development reduces a member bank's cost of holding reserves, but does not eliminate it because alternative assets yield a higher return. Unless regulated directly by the clearinghouse, a bank's reserve ratio is determined by precautionary liquidity considerations depending mainly on the volume and volatility of net clearings and the clearinghouse penalty for reserve deficiency (see Baltensperger 1980, 4-9; Santomero 1984, 584-6).

Once established, a clearinghouse may serve several purposes beyond the economical exchange and settlement of interbank obligations. It can become, in the words of James G. Cannon (1908, 97), "a medium for united action among the banks in ways that did not exist even in the imagination of those who were instrumental in its inception." One task the clearinghouse may take on is to serve as a credit information bureau for its members; by pooling their records, banks can learn whether loan applicants have had bad debts in the past or are overextended to other banks at present, and can then take appropriate precautions (Cannon 1910, 135). Through the clearinghouse banks can also share information concerning bounced checks, forgeries, and the like.

The clearinghouse may also police the soundness of each member bank in order to assure the other member banks that notes and deposits are safe to accept for clearing. As part of this function, banks may be required to furnish financial statements and may have their books audited by clearinghouse examiners. The Chicago clearinghouse insisted on statements as early as 1867, and in 1876 gained the right to carry out comprehensive examinations whenever desired, to determine any member's financial condition (James 1938, 372-3, 499). Regular examinations began in 1906 (Cannon 1910, 138-9). Other clearinghouses, such as the Suffolk Bank and the Edinburgh clearinghouse, took their bearings mainly from the trends of members' clearing balances and traditional canons of sound banking practice. Those two clearinghouses enjoyed such high repute as certifying agencies that to be taken off their lists of members in good standing meant a serious loss in reputation and hence business for an offending bank (Trivoli 1979, 20; Graham 1911, 59).

It is possible that a clearinghouse may attempt to organize collusive agreements on interest rates, exchange rates, and fee schedules for its members. However, rates inconsistent with the results of competition would tend to break down under unregulated conditions, for the standard reason that secretly underbidding a cartel has concentrated benefits and largely external costs. A clear example of this comes from Scottish experience (Checkland 1975, 391-427). The Edinburgh banks set up a committee in 1828 to set borrowing and lending rates. The Glasgow banks

THE EVOLUTION OF A FREE BANKING SYSTEM

joined a new version of the committee in 1836, at which time it represented the preponderance of Scottish banks in number and in total assets. Though not a clearinghouse association itself, the committee had much the same membership as the Edinburgh clearinghouse. In spite of repeated formal agreements, the committee could not hold members to its recommended interest rates. Not until after entry to the industry was closed in 1844 did the agreements become at all effective.

Perhaps the most interesting of all the roles a clearinghouse may perform is to assist its members in times of crisis (see Canon 1910, 24). If a bank or group of banks is temporarily unable to pay its clearing balances, or if it experiences a run on its commodity money reserves, the clearinghouse can serve as a medium through which more liquid banks lend to less liquid ones. It provides the framework for an intermittent, short-term credit market similar to the continuous federal funds market from which reserve-deficient American banks presently borrow. Another possible emergency function of clearinghouses is note issue. This function is called for when member banks are artificially restricted from issuing, as for example US banks were by the bond collateral requirements of the National Banking Acts, so that the banks are not able independently to fulfill all of their depositors' requests for hand-to-hand means of payment. Currency shortages occurred frequently in the United States during the second half of the nineteenth century, and clearinghouses helped to fill the void caused by deficient note issues of the National Banks.⁹

THE MATURE FREE BANKING SYSTEM

We have now reached the stage of mature development of a stylized free banking system, insofar as historical evidence illuminates its likely structural and operational characteristics. Evidence on industry structure from Scotland, Canada, Sweden, and elsewhere indicates that unregulated development does not produce natural monopoly, but rather an industry consisting of numerous competing banking firms, most having widespread branches, all of which are joined through one or more clearinghouses. In Scotland there were nineteen banks of issue in 1844, the final year of free entry. The largest four banks supplied 46.7 per cent of the note circulation. In addition to their head offices the banks had 363 branch offices, 43.5 per cent of which were owned by the largest (measured again by note issue) four banks.¹⁰

The banks in the mature system issue inside money in the shape of paper notes and demand deposit accounts (checkable either by paper or electronic means) that circulate routinely at par. Banks may also issue redeemable token coins, more durable but lighter and cheaper, to take the place of full-bodied coins as small change. Each bank's notes and tokens bear distinct brand name identification marks and are issued in the denominations the

public is most willing to hold. Because of the computational costs that would be involved in each transfer, interest is not likely to accrue on commonly used denominations of banknotes or tokens, contrary to the hypothesis of Neil Wallace (1983) that all currency would bear interest under *laissez-faire*.¹¹ Checkable accounts, however, provide a competitive yield reflecting rates available on interest-earning assets issued outside the banking system.

Checkable bank accounts are most familiarly structured as demand deposits, i.e., liabilities having a predetermined payoff payable on demand. An important reason for this structure is that historically a debt contract has been easier for the depositor to monitor and enforce than an equity contract which ties the account's payoff to the performance of a costly-to-observe asset portfolio. The predetermined payoff feature, however, raises the possibility of insolvency and consequently of a run on the bank if depositors fear that the last in line will receive less than a full payoff. One method of forestalling runs that may prevail in an unregulated banking system is the advertised holding of a large equity cushion, either on the bank's books or off them in the form of extended liability for bank shareholders. If this method were inadequate to assure depositors, banks might provide an alternative solution by linking checkability to equity or mutual-fund type accounts with postdetermined rather than predetermined payoffs. The obstacles to such accounts (asset-monitoring and enforcement costs) have been eroded over the centuries by the emergence of easy-to-observe assets, namely publicly traded securities. Insolvency is ruled out for a balance sheet without debt liabilities, and the incentive to redeem ahead of other account holders is eliminated. An institution that linked checkability to equity accounts would operate like a contemporary money-market mutual fund, except that it would be directly tied into the clearing system (rather than having to clear via a deposit bank). Its optimal reserve holdings would be determined in the same way as those of a standard bank.

The assets of unregulated banks would presumably include short-term commercial paper, bonds of corporations and government agencies, and loans on various types of collateral. Without particular information on the assets available in the economy, the structure of asset portfolios cannot be characterized in detail, except to say that the banks presumably strive to maximize the present value of their interest earnings, net of operating and liquidity costs, discounted at risk-adjusted rates. The declining probability of larger liquidity needs, and the trade-off at the margin between liquidity and interest yield, suggest a spectrum of assets ranging from perfectly liquid reserves, to highly liquid interest-earning investments (these constitute a "secondary reserve"), to less liquid higher-earning assets. Thus far, because the focus has been on monetary arrangements, the only bank liabilities discussed have been notes and checking accounts. Unregulated banks would almost certainly diversify on the liability side by offering a

tion of bullion and token fractional coins, issue of credit cards, and management of mutual funds. Such banks would fulfill the contemporary ideal of the "financial supermarket," with the additional feature of issuing banknotes.

Commodity money seldom if ever appears in circulation in the mature system, virtually all of it (outside numismatic collections) having been offered to the banks in exchange for inside money. Some commodity money will continue to be held by clearinghouses so long as it is the ultimate settlement asset among them. In the limit, if inter-clearinghouse settlements were made entirely with other assets (perhaps claims on a super-clearinghouse which itself holds negligible commodity money), and if the public were completely weaned from holding commodity money, the active demand for the old-fashioned money commodity would be wholly nonmonetary. The flow supply formerly sent to the mints would be devoted to industrial and other uses. Markets for those uses would determine the relative price of the commodity. The purchasing power of monetary instruments would continue to be fixed by the holder's contractual right (even if never exercised) to redeem them for physically specified quantities of the money commodity. The problem of meeting any significant redemption request (e.g., a "run" on a bank) could be contractually handled, as it was historically during note-duelling episodes, by invoking an "option clause" that allows the bank a specified period of time to gather the necessary commodity money while compensating the redeeming party for the delay. The clause need not (and historically did not) impair the par circulation of bank liabilities.

This picture of an unregulated banking system differs significantly in its institutional features from the visions presented in some of the recent literature on competitive payments systems. The system described here has assets fitting standard definitions of money. Banks and clearinghouses hold (except in the limit), and are contractually obligated to provide at request, high-powered reserve money (commodity money or deposits at the clearinghouse), and they issue debt liabilities (inside money) with which payments are generally made. These features contrast with the situation envisioned by Black (1970) and Fama (1980), in which "banks" hold no reserve assets and the payments mechanism operates by transferring equities or mutual fund shares unlinked to any money.

Bank reserves do not disappear in the evolution of a free banking system,

Banks, on the other hand, have a competitive incentive to redeem one another's liabilities regularly. So long as net clearing balances have a positive probability of being nonzero, reserves will continue to be held. In a system without reserve money it is not clear what would be used to settle clearing balances. In an evolved system, the scarcity of the money commodity and the costliness of holding reserves moreover serve to pin down the price level and to limit the quantity of inside money. In a moneyless system it is not clear what forces limit the expansion of payment media nor what pins down the price level. Nor are these things clear, at the other extreme, in a model of multiple competing fiat monies.¹²

Our analysis indicates that commodity-based money would persist in the absence of intervention, for the reason that the supreme salability of the particular money good is self-reinforcing. This result contradicts recent views (see Black 1970; Fama 1980; Greenfield and Yeager 1983; Yeager 1985) that associate complete deregulation with the replacement of monetary exchange by a sophisticated form of barter. (To be sure, Greenfield and Yeager recognize that their system would be unlikely to emerge without deliberate action by government, particularly given a government-dominated monetary system as the starting point.) In a commodity-based-money economy, prices are stated in terms of a unit of the money commodity, so the question of using an abstract unit of account does not arise as it does in a sophisticated barter setting.¹³ Even if actual commodity money were to disappear from reserves and circulation, the media of exchange would not be "divorced" from the commodity unit of account; they would be linked by redeemability contracts. We can see no force severing this link. Contrary to Woolsey (1985), the renunciation of commodity redemption obligations is not compelled by economization of reserves. Thus we find no basis for the spontaneous emergence of a multicommodity monetary standard or of any pure fiat monetary standard, such as contemplated in works by Hall (1982), Woolsey (1984), Klein (1974), and Hayek (1978). In short, unregulated banking would be much less radically unconventional, and much more akin to existing financial institutions than recent literature on the topic suggests.

One important contemporary financial institution is nonetheless missing from our account, namely the central bank. We find no market forces leading to the spontaneous emergence of a central bank, in contrast to the view of Charles Goodhart. (For this discussion a central bank is closely enough defined, following Goodhart (1985, 3–8), as an agency with two related powers: monetary policy, and external regulation of the banking system.) Goodhart (1985, 76) argues that the development of a central bank is "natural" because "the natural process of centralization of inter-bank deposits with leading commercial banks tends toward the development of a banks' club" which then needs an independent arbiter. But even on his own account the forces that historically promoted centralized interbank

THE EVOLUTION OF A FREE BANKING SYSTEM

deposits were *not* “natural” in any *laissez-faire* sense. They stemmed crucially from legal restrictions, particularly the awarding of a monopoly of note issue or the suppression of branch banking. Where no legislation inhibits the growth of branched banking firms with direct access to investment markets in the economy’s financial center, and able to issue their own notes, it is not at all apparent that profit seeking compels any significant interbank depositing of reserves. Walter Bagehot (1873, 66–8) argued persuasively that “the natural system – that which would have sprung up if Government had let banking alone – is that of many banks of equal or not altogether unequal size” and that in such a system no bank “gets so much before the others that the others voluntarily place their reserves in its keeping.” None of the relevant historical cases (Scotland, Canada, Sweden) shows any significant tendency toward interbank deposits.

We have seen that reserves do tend to centralize, on the other hand, in the clearinghouses. And clearinghouses, as Gorton (1985a, 277, 283; 1985b, 274) has recently emphasized, may take on functions that are today associated with national central banks: holding reserves for clearing purposes, establishing and policing safety and soundness standards for member banks, and managing panics should they arise. But these functional similarities should not be taken to indicate that clearinghouses have (or would have) freely evolved into central banks. The similarities instead reflect the pre-emption of clearinghouse functions by legally privileged banks or, particularly in the founding of the Federal Reserve System (Gorton 1985a, 277; Timberlake 1984), the deliberate nationalization of clearinghouse functions. Central banks have emerged from legalization contravening, not complementing, spontaneous market developments.¹⁴

NOTES

- 1 See for example Black (1970), Klein (1974), Hayek (1978), Fama (1980), Greenfield and Yeager (1983), Wallace (1983), White (1984b), O’Driscoll (1985a), and Yeager (1985).
- 2 We have each made normative evaluations of free banking elsewhere: Selgin (1988, chs. 8–10); White (1984a, ch. 5, 1984b).
- 3 See also Menger (1981, 260–2). The same view appears in Carlisle (1901, 5), and Ridgeway (1892, 47). A more recent version of Menger’s theory is Jones (1976). For a secondary account of Menger’s theory see O’Driscoll (1985b).
- 4 See Menger (1981, 263–6); Ridgeway (1892, 6–11), and Burns (1927a, 286–8). On some alleged nonmetallic monies of primitive peoples see Quiggin (1963).
- 5 See Usher (1943), de Roover (1974, chs. 4–5), and Lopez (1979).
- 6 On the historical development of banknotes and checks in Europe see Usher (1943, 7–8, 23).
- 7 See White (1984a, 84–5) for nineteenth century views on geographic diseconomies in note circulation.
- 8 An example of the explicit adoption of “tit for tat” by an exhausted note-duelling bank is given by Munn (1981, 24).

THE NATURE OF FREE BANKING

- 9 See Cannon (1908), Andrew (1908), Smith (1936), Timberlake (1984), and Gorton (1985a).
- 10 These figures are based on data in White (1984a, 37). A recent econometric study of economies of scale in banking is Benston, Hanweck, and Humphrey (1982).
- 11 See White (1984a, 8–9, 1987).
- 12 Taub (1985) has shown that a dynamic inconsistency facing issuers in Klein's (1974) model will lead them to hyperinflate. Cf. Chapter 1.
- 13 This point is emphasized by White (1984c). For additional criticism of the Black–Fama–Yeager literature see O'Driscoll (1985a), Hoover (1985a), and McCallum (1984).
- 14 On the appearance of central banks in several nations see Smith (1936); on Canada in particular see Bordo and Redish (1987).

REFERENCES

- Andrew, A. Piatt. "Substitutes for Cash in the Panic of 1907." *Quarterly Journal of Economics*, August 1908, 497–516.
- Axelrod, Robert. *The Evolution of Cooperation*. New York: Basic Books, 1984.
- Bagehot, Walter. *Lombard Street. A Description of the Money Market*. London: Henry S. King, 1873.
- Baltensperger, Ernst. "Alternative Approaches to the Theory of the Banking Firm." *Journal of Monetary Economics*, January 1980, 1–37.
- Benston, George J., Gerald A. Hanweck, and David B. Humphrey. "Scale Economies in Banking: A Restructuring and Reassessment." *Journal of Money, Credit, and Banking*, November 1982, 435–54.
- Bisschop, W. R. *The Rise of the London Money Market, 1640–1826*. London: P. S. King & Son, 1910.
- Black, Fischer. "Banking and Interest Rates in a World Without Money: The Effects of Uncontrolled Banking." *Journal of Bank Research*, Autumn 1970, 9–20.
- Bordo, Michael and Angela Redish. "Why Did the Bank of Canada Emerge in 1935?" *Journal of Economic History*, June 1987, 405–17.
- Burns, A. R. "Early Stages in the Development of Money and Coins," in *London Essays in Economics in Honour of Edwin Cannan*, edited by T. E. Gregory and Hugh Dalton. London: George Routledge & Sons, 1927a.
- . *Money and Monetary Policy in Early Times*. New York: Alfred E. Knopf, 1927b.
- Cannon, James G. "Clearing Houses and the Currency," in *The Currency Problem and the Present Financial Situation*. New York: Columbia University Press, 1908.
- . *Clearing Houses*. Washington, DC: Government Printing Office, 1910.
- Carlisle, William. *The Evolution of Modern Money*. London: Macmillan, 1901.
- Checkland, S. G. *Scottish Banking: A History, 1695–1973*. Glasgow: Collins, 1975.
- de Roover, Raymond. *Business, Banking, and Economic Thought in Late Medieval and Early Modern Europe*, edited by Julius Kirshner. Chicago: University of Chicago Press, 1974.
- Fama, Eugene F. "Banking in the Theory of Finance." *Journal of Monetary Economics*, January 1980, 39–57.
- . "Financial Intermediation and Price Level Control." *Journal of Monetary Economics*, July 1983, 7–28.

THE EVOLUTION OF A FREE BANKING SYSTEM

- Gibbons, J. S. *The Banks of New York, Their Dealers, the Clearing House, and the Panic of 1857*. New York: D. Appleton Co., 1858.
- Goodhart, Charles. *The Evolution of Central Banks: A Natural Development?* London: Suntory-Toyota International Centre for Economics and Related Disciplines/London School of Economics and Political Science, 1985.
- Gorton, Gary. "Clearinghouses and the Origin of Central Banking in the United States." *Journal of Economic History*, June 1985a, 277–83.
- . "Banking Theory and Free Banking History: A Review Essay." *Journal of Monetary Economics*, September 1985b, 267–76.
- Graham, William. *The One Pound Note in the History of Banking in Great Britain*, 2nd edn. Edinburgh: James Thin, 1911.
- Greenfield, Robert L. and Leland B. Yeager. "A Laissez Faire Approach to Monetary Stability." *Journal of Money, Credit, and Banking*, August 1983, 302–15.
- Hall, Robert. "Explorations in the Gold Standard and Related Policies for Stabilizing the Dollar," in *Inflation Causes and Effects*, edited by Robert Hall, Chicago: University of Chicago Press for the National Bureau of Economic Research, 1982.
- Hayek, F. A. *Denationalisation of Money*, 2nd edn. London: Institute of Economic Affairs, 1978.
- Hicks, John. "The Two Triads, Lecture I," in *Critical Essays in Monetary Theory*. Oxford: Clarendon Press, 1967.
- Hoover, Kevin D. "Causality and Invariance in the Money Supply Process." Doctoral dissertation, Oxford University, 1985.
- James, F. Cyril. *The Growth of Chicago Banks*. New York: Harper & Brothers, 1938.
- Jones, Robert. "The Origin and Development of Media of Exchange." *Journal of Political Economy*, November 1976, 757–75.
- Klein, Benjamin. "The Competitive Supply of Money." *Journal of Money, Credit, and Banking*, November 1974, 423–53.
- Lake, Wilfred S. "The End of the Suffolk System." *Journal of Economic History*, November 1947, 183–207.
- Leslie, J. O. *The Note Exchange and Clearing House Systems*. Edinburgh: William Blackwood, 1950.
- Lopez, Robert S. "The Dawn of Medieval Banking," in *The Dawn of Modern Banking*. New Haven, CT: Yale University Press, 1979.
- McCallum, Bennett T. "Bank Deregulation, Accounting Systems of Exchange, and the Unit of Account: A Critical Review." Carnegie-Rochester Conference Series on Public Policy, Autumn 1984, 13–45.
- Melitz, Jacques. *Primitive and Modern Money*. Reading, MA: Addison-Wesley, 1974.
- Menger, Carl. "The Origin of Money." *Economic Journal*, June 1892, 239–55.
- . *Principles of Economics* [1871]. New York: New York University Press, 1981.
- Munn, Charles W. "The Origins of the Scottish Note Exchange." *Three Banks Review* 107, 1975, 45–60.
- . *The Scottish Provincial Banking Companies, 1747–1864*. Edinburgh: John Donald, 1981.
- O'Driscoll, Gerald P., Jr. "Money in a Deregulated Financial System." *Economic Review*, Federal Reserve Bank of Dallas, May 1985a, 1–12.
- . "Money: Menger's Evolutionary Theory." Unpublished manuscript, Federal Reserve Bank of Dallas, 1985b.

THE NATURE OF FREE BANKING

- Powell, Ellis T. *The Evolution of the Money Market, 1385-1915*. New York: Augustus M. Kelley, 1966.
- Quiggin, A. Hingston. *A Survey of Primitive Money: The Beginning of Currency*. London: Methuen, 1963.
- Richards, R. D. *The Early History of Banking in England*. New York: Augustus M. Kelley, 1965.
- Ridgeway, William. *The Origin of Metallic Currency and Weight Standards*. Cambridge: Cambridge University Press, 1892.
- Santomero, Anthony M. "Modeling the Banking Firm: A Survey." *Journal of Money, Credit, and Banking*, November 1984, 576-602.
- Selgin, George A. *The Theory of Free Banking*. Totowa, NJ: Rowman and Littlefield, 1988.
- Smith, Vera C. *The Rationale of Central Banking*. London: P. S. King, 1936.
- Taub, Bart. "Private Fiat Money with Many Suppliers." *Journal of Monetary Economics*, September 1985, 195-208.
- Timberlake, R. H. "The Central Banking Role of Clearing-House Associations." *Journal of Money, Credit, and Banking*, February 1984, 1-15.
- Trivoli, George. *The Suffolk Bank: A Study of a Free-Enterprise Clearing System*. London: Adam Smith Institute, 1979.
- Usher, Abbott Payson. *The Early History of Deposit Banking in Mediterranean Europe*. Cambridge, MA: Harvard University Press, 1943.
- Wallace, Neil. "A Legal Restrictions Theory of the Demand for 'Money' and the Role of Monetary Policy." *Federal Reserve Bank of Minneapolis Quarterly Review*, Winter 1983, 1-7.
- White, Lawrence H. *Free Banking in Britain: Theory, Experience, and Debate, 1800-1845*. Cambridge: Cambridge University Press, 1984a.
- . "Free Banking as an Alternative Monetary System," in *Money in Crisis*, edited by Barry N. Siegel. San Francisco: Pacific Institute, 1984b.
- . "Competitive Payments Systems and the Unit of Account." *American Economic Review*, September 1984c, 699-712.
- . "Accounting for Non-Interest-Bearing Currency: A Critique of the 'Legal Restrictions' Theory of Money." *Journal of Money, Credit, and Banking*, November 1987, 448-56.
- Williams, Jeffrey C. "Fractional Reserve Banking in Grain." *Journal of Money, Credit, and Banking*, November 1984, 488-96.
- Withers, Hartley. *The Meaning of Money*. London: John Murray, 1920.
- Woolsey, Warren W. "The Multiple Standard and the Means of Exchange." Unpublished manuscript, Talladega College, Talladega, AL, 1984.
- . "Competitive Payments Systems: Comment." Unpublished manuscript, Talladega College, 1985.
- Yeager, Leland B. "Deregulation and Monetary Reform." *American Economic Review*, May 1985, 103-7.