1

The Evolution of Market Monetary Institutions

In the beginning, goods were bartered directly for other goods. A theoretical account of the evolution of monetary institutions naturally begins with an attempt to explain how the earliest money emerged from a non-monetary or barter economy. "Money" here, following standard usage among economists, means a *commonly accepted medium of exchange.* A "medium of exchange" means a good that people acquire through trade with the intention of trading away later (rather than consuming for its own sake or using up in a production process). "Commonly accepted" means that the money good is routinely offered and taken in trade for other goods, and so appears on one side of nearly every transaction. The theory that follows aims to explain why, and how, some good should acquire these characteristics.[1]

The Mystery of Money

The Austrian economist, Carl Menger (1892), developed the classic explanation of the origin of money. Menger showed how money can emerge from barter without anyone *inventing* it, or to use Adam Smith's phrase, "as if by an invisible hand." In his account, money emerges through a series of steps, each based on self-seeking actions by individual traders, without the resulting social order (monetary exchange) being part of anyone's intention. This

[1] For a further teasing out of these defining terms see White (1989, ch. 11). There I used the modifier "generally accepted"; here, following Wärneryd (1990), I use "commonly accepted" in the same sense. The present chapter draws on Selgin and White (1987) and White (1989, ch. 9). For complementary accounts of these issues, see Glasner (1989, ch. 1) and Dowd (1996, ch. 1).

is a satisfying mode of explanation because it does not require heroic assumptions about the knowledge possessed by any trader.

A number of writers before Menger expressed the idea that money was an undesigned or spontaneously emerged institution. Among them are Adam Smith, the French economists Etienne de Condillac and Destutt de Tracy, and the British monetary pamphleteers, Thomas Hodgskin and Samuel Bailey. Menger was certainly aware of Smith's writings, though he does not cite Smith in this context. However, none of these earlier writers spelled out the emergence of money step by step. The typical modern textbook discussion of the origin of money is plainly inadequate.[2] It lists the problems of barter exchange, and shows that monetary exchange overcomes these problems. A prototype can be found in Aristotle (*Nichomachean Ethics*, Book 5): "All the things which we exchange need to be comparable. This need led to the invention of money to serve as a medium giving value to every thing." Unfortunately, the simple contrast between problems and solution does not explain how the solution (money) was arrived at, any more than a list of the advantages of standard time zones would explain how they came about. One is left with the impression that barterers, one morning, suddenly became alert to the benefits of monetary exchange, and, by that afternoon, were busy using some good as money. In one version of the story, a wise head of state introduced the idea that a certain commodity was to be sanctioned as a general medium of exchange.

Taken seriously as a theory of the origin of money, this account would suggest that the idea of money was fully grasped before money existed. Money would be an invention, like the telephone, which existed in someone's mind before a prototype was produced. In fact, money is not a product of technological advance brought forth by a single mind or a research laboratory. This is evident from the fact that gold dust or salt, used as money, is not technologically different from gold dust or salt, not used as money. What transforms gold dust or salt into a money is not some physical change, but rather the development of a *social convention* concerning the use of that good. The use of any particular item as money is a social convention, in the same sense that the use of particular utterances or gestures to communicate particular ideas is a social convention. Each of us (in an English-speaking group) calls a certain fruit an "apple" because that is what everyone around us calls it, and we wish to communicate with them. Likewise, each of us uses item x as a medium of exchange because nearly all others in our society do, and we wish to trade with them.

A money could not spring forth full-blown from barter unless people throughout a society simultaneously arrived at the idea of using x as a medium of exchange, *and* each person knew that he could count on others to

[2] Notable exceptions are McCulloch (1982) and Goodhart (1989b).

do so too. Such a scenario begs too many questions. It invokes the realization of money in the attempt to explain how money was realized. It attributes knowledge of the benefits of money to people who would not have such knowledge in a barter economy.

Menger begins by emphasizing the "mystery" of money: why is everyone willing to trade truly useful goods and services for mere tokens? In Menger's day (a century ago), these tokens were otherwise practically useless disks of gold and silver, or slips of paper (banknotes) representing claims to such disks. Today, the mystery is even greater, as the tokens are otherwise completely useless disks of cupro-nickel and slips of paper interchangeable with them.[3]

Menger's approach does not apply only to commodity money, though it was originally framed to explain such money. It emphasizes that the use of a commodity money has a "conventional" aspect, the convention being one that develops through a historical process. By extension, the use of a fiat money rests on the prior development of a commodity money convention, because fiat money is launched by suspending the redeemability of claims to a commodity money. However, we are getting ahead of the story.

It is worthwhile restating Menger's theory in detail for several reasons. Our immediate interest, here, is its usefulness in explaining the origin of money. Later in the book, we will return to the theory because it has implications for the viability of projects to establish a new money, or a payments system without money. The theory also draws out certain "essential features" of money that have implications for the macroeconomic properties of a monetary economy (Yeager 1968). Finally, the theory holds a general interest to students of the social sciences because it provides a paradigmatic example of an invisible-hand explanation of a social institution.[4]

Menger's Theory Restated

A simple barter economy faces each trader with the problem of finding a trading partner with preferences and endowments reciprocal to his own. (This has come to be known as the problem of finding a "mutual coincidence of

[3] In recent years, a number of monetary economists have offered non-evolutionary models of money as solutions to the mystery of a positive value being accorded to "intrinsically useless" and inconvertible fiat money: in particular, the overlapping generations model (Wallace 1980) and search-theoretic models (Kiyotaki and Wright 1989, Ritter 1995). Menger's solution is different, and is less subject to the cogent criticisms made of the overlapping generations model (Tobin 1980, McCallum 1983), of other general equilibrium models of money (Bryant and Wallace 1980), and of search-theoretic models (Selgin 1997b). Of course, it is subject to other criticisms.

[4] It has been cited as such by Nozick (1974, p. 18), though Nozick actually cites a restatement of Menger's theory by von Mises (1980).

wants.") For Alanis to trade the asparagus she brings to market for the bacon she prefers to take home and consume, via a direct pairwise exchange, she must find some other trader ("*B*") who both has what Alanis wants (bacon), and wants what Alanis has (asparagus). It may be difficult or even impossible to find such a match, even when a unanimously preferred reallocation of goods could be arranged in another way. McCulloch (1982, ch. 1) offers a simple example: imagine three individuals and three indivisible goods, where *A* has good 1 and prefers good 2 (but not good 3) to good 1, *B* has good 2 and prefers only good 3, and *C* has good 3 and prefers only good 1. Clearly, all are better off if good 1 goes to C, good 2 goes to A, and good 3 goes to B, but there is no pairwise exchange that makes both traders better off. More generally, even where a pairwise trading partner *could* be found, it may be difficult and time-consuming to find that trader among the many in the marketplace.

A trader who is frustrated by her inability to find a rare or non-existent matched trading partner need neither continue the effort fruitlessly nor give up and go home. There is an alternative. Consider the three-agent case just described. Suppose that each pair of individuals has met, and has quickly discovered that only one party wants to make each possible pairwise trade. Alanis has discovered that Bjork, who is selling the good that Alanis prefers (bacon), does not prefer what Alanis has to offer (asparagus). In this situation, it would not take too much cleverness on Alanis's part to ask Bjork what good Bjork *would* prefer. Learning that Bjork prefers cabbage, which Coolio has offered to Alanis in exchange for asparagus, Alanis will be led by self-interest to trade with Coolio, even though Alanis does not want to consume cabbage. Alanis will then be in a position to make an offer for Bjork's bacon that Bjork will accept.

The general point illustrated by this example is that, potentially, a barterer can economically achieve a preferred holding of goods by exchanging her initial endowment for some good which can then be turned around and exchanged for the good(s) she ultimately wants to consume. This practice is known as *indirect exchange*, in contradistinction to the *direct exchange* of simple barter. In the example, individual *A* has used good 3 as a vehicle for indirect exchange or, as it is usually put, as a *medium of exchange.*

Now consider a larger barter market, such as a trade fair, with anonymous traders selling many goods. (As noted below, such a fair would not historically have been found in a barter economy, because specialization and trade could not develop far where trade remained so difficult to accomplish.) To keep the discussion simple, assume that each trader still arrives endowed with a single indivisible good and desires to take home some one other good, though several traders may now be selling each good. As before, every trader besides Alanis is trying to use direct exchange, and will agree to trade only for the one good she wishes to consume. Alanis,

again coming to market with asparagus, will find that her trading problem is now more difficult. With a larger number of traders, it may take more time to discover a trader *B* who is selling the bacon that Alanis wants to buy. The probability that this *B* wants to take home asparagus is smaller than before. Once Alanis learns what good *j* this trader *B will* agree to accept in exchange, it will take more time to find which (if any) among the sellers of *j* will accept asparagus in exchange. In this setting, it can easily be the case that Alanis acquires the desired bacon most economically neither by direct exchange, nor by waiting until she meets a seller of bacon to learn what goods can potentially be used as a medium of exchange. Instead, Alanis's best trading strategy is to exchange her asparagus for a good *k* which any seller of bacon (and, for that matter, of good *j*) that Alanis may happen to meet is *relatively likely* to accept.

Here Menger introduces the concept that different goods have different *degrees of marketability*. Marketability is a "non-Walrasian" concept: in a Walrasian general equilibrium model, with costlessly coordinated trade, and with a single price at which a good may be either bought *or* sold, every good is perfectly marketable. In a world of costly trade, it takes some amount of time, effort, and expense to sell for a good price. (Anyone who has ever tried to sell a used car knows this.) A more highly marketable good is a good that is easier (less costly) to sell for a good price. A "good price" here means a price close to the best price that could potentially be found with full information on both sides of the market. (Menger calls this best price, an "economic price.")[5]

A perceptive barterer will exchange her initial endowment goods for more highly marketable goods, which can then easily be exchanged for the goods he or she wants to consume. She wants to maximize her expected gains from trade, which obviously depend on the prices at which trade takes place, net of the costs of finding trading partners (costs of search), and of the costs of consummating trade (costs of transportation, contracting, and the like). Indirect exchange requires two trades, instead of one. It is therefore more likely to be advantageous to the extent that:

1 the good to be used as a medium of exchange is more widely consumed, and traded, than the endowment good, and hence trading partners offering a good price in it, and for it, are easier to find; and
2 the costs of buying, holding, and reselling it (costs of contracting, spoilage, and transportation) are relatively small.

[5] Because marketability has at least three dimensions (time cost, other selling costs, and percentage of economic price realized), it may not always be possible to rank the marketability of different goods unambiguously.

Where indirect exchange is advantageous, it pays a trader to accumulate an inventory of highly marketable items for use as media of exchange.[6] Having highly marketable items on hand allows a trader to find good buying prices more easily for the things she wants to consume.

Other alert traders in the market, facing the same situation, will adopt the same strategy of indirect exchange. Menger notes that some individuals may not catch on to the advantages of indirect exchange immediately or on their own. Eventually though, they will notice the success enjoyed by those who are trading their produce for a medium of exchange rather than persisting in attempts at direct exchange. They are then likely to imitate the practice of using indirect exchange.

Once many individuals are using indirect exchange, the stage is set for social convergence toward a *common* medium of exchange. One perceptive trader, say *M*, will learn from experience which commodities are *most* marketable, and best suited for use as media of exchange. The knowledge that he can unload them easily will lead him to accept these commodities all the more readily, and in preference to other commodities. *M*'s greater acceptance of a good *k* incrementally reinforces its usefulness as a medium of exchange for other traders, *A* . . . *L* and *N* . . . *Z*, because they can count on one more place to spend it. Its marketability for them has increased. They may learn of good *k*'s improved suitability as a medium of exchange, either through communication, or from trial-and-error experience, or as a last resort by imitation of the successful traders. Traders *N*, *O*, and the others will then accept good *k* more readily, just as *M* did earlier. Again, each trader who does so reinforces its usefulness for the others. With every trader preferring more marketable to less marketable media of exchange, ultimately one good (or at most a few, covering different sets of transactions) is elevated to the status of being *commonly* or *generally* or *routinely* accepted as a medium of exchange. It becomes money.

This theory is not meant to suggest that extensive specialization and market trade historically antedated the emergence of money. On the contrary, it helps to explain why specialization and trade developed simultaneously with money, a fact Leijonhufvud (1981, pp. 229–30) has emphasized. Pre-monetary communities were basically autarkic (Dingle 1988). Direct exchange is so difficult that the scope of specialized production "for the market" is limited by the

[6] This was pointed out long ago by Adam Smith (1981, pp. 37–8): "In order to avoid the inconveniency of such situations [in which the seller of a desired good does not want the produce the would-be buyer has to offer], every prudent man in every period of society, after the first establishment of the division of labor, must naturally have endeavoured to manage his affairs in such a manner, as to have at all times by him, besides the peculiar produce of his own industry, a certain quantity of some one commodity or other, such as he imagined few people would be likely to refuse in exchange for the produce of their industry."

scope of indirect exchange. Thus Adam Smith's dictum, that the division of labor is limited by the extent of the market, may be supplemented by the observation that the extent of the market is limited by the extent of money's use.

Some Implications of the Theory

The Mengerian theory helps us to understand the important characteristics of a monetized economy that are the result of these evolutionary origins of money. Menger emphasized the following three.

1 Everyone in a monetary economy routinely accepts money, and routinely attempts to trade output or endowment goods for money before acquiring consumption goods.
2 The ability to purchase goods at "the going rate" with money is not at all doubtful, even in anticipation of dealing with an anonymous seller. There is virtually no risk of meeting a seller who refuses to accept money, or accepts it only at a discount.
3 Sellers are reluctant to accept goods of lesser marketability than money, with the result that the marketability of the money good is discontinuously greater than that of any other good. A buyer (with money) has markedly less difficulty trading at close to economical prices than does a seller (of a non-money good). As Menger notes, being forced to sell on short notice imposes much more of a burden than being forced to buy on short notice.

The theory also establishes that *no collective decision* or legislative act is necessary for money to emerge. Menger emphasized this point with respect to the defining characteristic of money: its general acceptance as a medium of exchange. Money did not originate from, or fail to perform its medium-of-exchange function fully until endorsed by, the legal decrees of rulers.

By extending Menger's theory, we can see that the role of money as a "unit of account" also arises spontaneously. The "unit of account" means some definite quantity of a good used as a pricing and accounting unit. Strictly speaking, as Jurg Niehans (1978) has pointed out, it is not proper to say that *money* is a unit of account, because money as such is not a unit. Money is rather a *medium* of account. The *unit* of account is a specific quantity of the good constituting the medium of account. For example, silver may be the money and medium of account, while the "ducat" (defined as so many grams of standard-fineness silver) is the unit of account.[7]

[7] The ducato d'argento of Venice, 1201–1355, was 96.5% fine and weighed 2.18 grams.

The medium of account in an economy will naturally coincide with its commonly accepted medium of exchange or money. A seller pursues his self-interest by posting prices in terms of the good or goods he is routinely willing to accept in exchange. If this seller were to post prices in terms of some other good, he would incur the extra trouble, for himself and for his customers, of keeping track of and figuring in the current exchange rate between the pricing medium and the payment medium he is willing to accept. An accountant could not calculate profit and loss as easily, or as clearly, were she to keep books in units of a commodity other than the commonly accepted medium of exchange in which the income actually accrues, liabilities come due, transactions balances are held, and for which other assets can most readily be exchanged.

Accordingly, some common unit of money naturally becomes the unit of account. It may be an early popular coin (e.g. the Spanish dollar), a simple bullion weight (e.g. the "pound sterling"), or a natural unit (one standard-sized cowry shell). No official proclamation is necessary to establish a standard unit of account, any more than to establish a standard building brick. Commercial practice can converge on a conventional unit, without any collective decision being taken, in much the way that it converges on a money commodity. Each seller will discover that he does best for himself by posting prices in the unit most popular with his potential trading partners. Court decisions can, and historically did, follow trade custom in deciding how much of what fineness of what metal would legally satisfy a contractual obligation to deliver a "ducat" (or whatever money unit). The courts did not *create* customary units.

There is no denying, of course, that governments often have played a role in pushing a new money or unit of account. The point is that no collective deliberation or action is needed for money to emerge fully, or was historically instrumental in its original emergence.[8]

Money only makes sense, given its origin in indirect exchange, in a world with certain features. There must be three or more traders, and at least two goods besides money, for otherwise only direct exchange is possible. There must be varying degrees of marketability for these goods, or some physical feature of other goods that makes them less suitable for shopping with; otherwise, there is no advantage to indirect exchange. These features are

[8] John Maynard Keynes (1935, pp. 4–5), even while asserting the relevance to the modern world of "the doctrine that money is peculiarly a creation of the State," recognized that the original establishment of a conventional unit of account preceded government involvement: "Thus the Age of Money had succeeded to the Age of Barter as soon as men had adopted a money-of-account. And the Age of . . . State Money was reached when the State claimed the right to declare what thing should answer as money to the current money-of-account – when it claimed the right not only to enforce the dictionary but also to write the dictionary."

not always present in abstract economic models purported to be models of monetary economies. In a world where all goods are perfectly and costlessly marketable (so that there are no economic barriers to barter), there is no rationale for a distinctive money. Barter with interest-bearing financial assets would dominate the use of any non-interest-bearing money.

From Simple Commodity Money to Coins

An evolutionary or neo-Mengerian perspective can help to explain the emergence of gold and silver as the predominant commodity monies in the world, and the later emergence of such monetary institutions as coinage and bank-issued paper money.

The earliest form of money, following Menger's account, must have been a useful commodity. A good must have acceptability in barter before it can acquire wider acceptability as a medium of exchange. It must have some usefulness as a commodity to be accepted in barter. Anthropological evidence indicates that the goods that became monies in several cultures originally had ornamental uses (Melitz 1974).[9] This was true of Pacific and African shell monies, North American wampum, and, also, gold and silver. Other primitive monies have been foodstuffs, like grain or salt.

The eventual predominance of gold and silver as money, over other commodities which early on would have had equally wide acceptability, can be explained by at least four (partly physical) characteristics that promoted their ready marketability and convenience (low usage costs) as media of exchange. These characteristics were a staple subject of discussion in money-and-banking texts during the era of metallic monetary standards.

1 Goods like livestock or tobacco, whose quality is variable and difficult to assess, are more troublesome to exchange than goods of *uniform* and *easily recognized quality*.[10] Pure gold and silver, as chemical elements, are absolutely uniform. The purity (fineness) of a particular piece of gold or silver can be tested at low cost by biting it, sounding it, or (with a bit more trouble) by assaying it. Traders were commonly experienced in these assessment methods in the past. As

[9] I used to add "or ceremonial uses," until I discovered that "ceremonial use" is the anthropologist's shorthand for "we have no idea what it was used for."

[10] Armen Alchian's account (1977) of "Why Money?" relies exclusively on low authentication costs for selecting which commodity will become money. See also King and Plosser (1986). Other things equal, this characteristic can be decisive but, more generally, it is only one characteristic among several that can play a part in promoting a commodity's use as a medium of exchange.

will be discussed below, coinage arose to relieve the difficulties created by the non-uniformity of gold and silver in rawer forms (nuggets or dust or ingots).

2 Gold and silver are *durable*, so that there are no extra carrying costs due to spoilage. The deterioration of goods like grain and olive oil makes them costly to hold in inventory. The possibility of deterioration also creates the above-mentioned problem of exchange being encumbered by the need for costly verification of the goods' current quality.

3 The precious metals are easily *divisible and fusible*, so that payment can be tailored to purchase size. Large pieces can easily be split into small pieces, and small pieces can be united to form larger pieces. This is not true of jewels or, certainly, of livestock.

4 Finally, gold and silver are *portable*, that is, have high ratios of value to bulk. Portability means a low cost of taking the medium of exchange from the site where it is acquired to the site where it is spent. Commodities like salt lost their suitability as media of exchange when their value per pound became too low. The copper money of seventeenth-century Sweden, a non-precious metallic money, was notoriously cumbersome. Individual pieces of copper "plate money" eventually weighed up to 20 kilograms (44 pounds). Strong young men had to be employed to carry the copper necessary to make an ordinary-sized commercial purchase. Finally, Swedes stopped using copper, except in the smallest transactions. A similar process may have promoted the historical dominance of gold over silver in international payments of large sums: it was less costly to send one boat laden with gold than to send fifteen laden with silver.[11]

The displacement of one money by another can follow the general Mengerian logic of a self-reinforcing convergence process. As individuals from two regions with different commodity monies come into contact and begin to trade, an entryway is created for the better of the two monies to spread to the other region. Traders on the margin, not only those physically adjacent to the border but also merchants who do a large fraction of their trade with users of the foreign money, will favor the foreign money, if it is markedly better in some of the four areas listed above. Merchants and border-dwellers will accept the favored money on somewhat better terms, and

[11] Fleming (1994) finds, however, that the general historical switch from silver to gold standards was not market-driven. It was, in fact, mainly due to the legal overvaluation of gold relative to silver by the governments of Britain and the US, which set Gresham's Law in motion (the legally overvalued or "bad" money drove out the legally undervalued or "good" money). Other nations deliberately followed suit in a sort of bandwagon effect.

can use it among themselves where, before, they used the local money. The margin can then spread: those who deal substantially with these merchants, and those who live adjacent to the areas adjacent to the border, can find it advantageous to be paid in the foreign money. Its sphere of acceptance can snowball, following the Mengerian logic, until a single money unites the two regions.

Coinage, the practice of fashioning monetary metal into standardized marked discs, though it involves technical advances and not merely the formation of a social convention, also developed in step-by-step fashion. Where nuggets or gold dust served as money, merchants had to assess weight and quality when accepting payment. It made sense for a merchant to mark a piece of assessed gold, so as to avoid the costs of re-assessment when paying the piece out later. Other traders who trusted this merchant could then also rely on his mark. To prevent the possibility of shaving off gold around the marked area, the piece could be covered with marks. Punching, stamping, and finally modern methods of minting developed as low-cost methods of fashioning reliably marked pieces of gold. Historical examples of these stages can be observed in the money of ancient Lydia (Burns 1927; Cribb 1986).

Mints arose spontaneously, then, to meet the demand for authentication services. With the development of coinage, the marketability of coined metal became discontinuously greater than that of uncoined metal (in this context, branded bars of bullion may be thought of as large coins). Gold miners found it much easier to spend coined than uncoined gold, and, therefore, were willing to pay for the service of minting their raw gold into coins. Numismatic publications indicate that more than twenty private gold and silver mints operated during the gold and silver rushes in nineteenth-century America (Kagin 1981), and one in Australia (McDonald 1987, p. 122).

In practice, governments have typically monopolized the coinage industry, but there are no signs that coinage is a natural monopoly. There are ample signs that governments have wanted to exercise monopoly over money production so as to reap the monopoly profits known as *seigniorage* (Selgin and White 1999). In a later chapter, we will consider in more detail both seigniorage in its medieval form, and seigniorage in its modern form of the profit from monopoly issue of fiat money.

Bank-issued Money

The next step to consider, in the evolution of monetary institutions, is the emergence of money issued by commercial banks. Full-bodied coins (and other types of full-bodied commodity money, like shells) originate outside of any commercial banking system. We may call them "outside" money,

whereas bank-issued money is "inside" money. Outside money is an asset for its holder but not a liability of, or financial claim against, anyone else. The media of exchange produced by a commercial bank, by contrast, are claims against it. A large literature attempts to explain why banks exist as intermediaries between savers and borrowers (Santomero 1984).[12] Our object here is, rather, to explain why banks participate in the payments system, by offering a logical evolutionary account of why and how claims against banks came to be used as money.

The earliest bank liabilities were claims to outside money deposited with bankers. Historical records indicate that bankers in medieval Italy began as money-changers, but by AD1200 had moved into accepting time and demand deposits (de Roover 1974a, 1974b). In a region of numerous city-states, each with its own distinct coinage, money-changers provided the service of trading local coins for the less spendable foreign coins brought by inbound merchants and other travelers, and of trading the reverse way with outbound travelers. A simple explanation of why money-changers became deposit-takers is that merchants found it easier to leave money with them "on account," to be called for when needed, rather than to take away domestic coin equal in value to the foreign coin tendered (or vice versa) on every occasion. Essentially, this means that the money-changers' vaults were being used for temporary safekeeping of coin. In this respect, the development of deposit banking in Italy was similar to its development in England where, according to numerous accounts, early deposits were taken by goldsmiths whose vaults provided safekeeping.

Bank deposits began to play a monetary role when they became a medium of exchange, that is, when transfer of deposit balances became an accepted method of payment among bank customers. The practice of deposit transfer evolved by steps. Where a bank provided safekeeping services, depositors no doubt discovered cases in which party Alice planned to withdraw coins from the vault and laboriously transfer them to party Bob, who in turn planned to lug them back to the same vault and redeposit them. At the end of the day, the coins were back where they started, Alice's deposit balance had been reduced, and Bob's balance had been enlarged by the same amount. Only a little imagination was needed for Alice and Bob to recognize that an easier method of accomplishing this result would be for them to meet in the banker's office (in the coin-lugging method, both had to

[12] An intermediary is an institution that issues financial claims (debt or equity) against itself, and uses the proceeds to acquire financial claims on other agents. Because it is irredeemable and not a financial claim, fiat money is outside rather than inside money, and an institution that issues it (typically a central "bank") is not, in that respect, acting as an intermediary. The text's distinction between outside and inside money is different from the one used by Gurley and Shaw (1960).

go to there anyway) and there persuade the banker simply to transfer the desired amount of deposit balances *on his books*. Alice and Bob thereby avoid physical lugging around of coins, which simply stay in the vault. Early banking documents, studied by de Roover (1974a; 1974b), record such three-way meetings among payer, payee, and banker to authorize deposit transfers.

Later developments made transfers still easier to accomplish. Written slips for authorizing transfers made it unnecessary for both parties to travel to the banker's office. (In a checking system, Alice hands Bob a check, and only Bob goes to the bank, to deposit it; in a "giro" system, only Alice goes to the bank, to authorize the transfer into Bob's account.) Today, we see the growing use of *electronic* funds transfer, that is, methods of authorizing deposit transfers using electronic messages (sent using a telephone, home computer, automatic teller machine, or debit card and point-of-sale terminal) in place of slips of paper. These methods do not change the nature of the payment system as one of deposit transfer. The "front end" of the deposit transfer is different from writing a check, but not the "back end" (what happens on the bank's balance sheet). Nor – despite excited predictions that the future holds "a world without money" – do they threaten the definition, or real existence, of money. The depositor's bank balance, not the transfer-authorization device (e.g. the check), is money.

In addition to deposits, bank-issued claims in currency form were important historically, and may soon become important again. *Banknotes* are bank-issued claims to outside money that are not in any customer's name, but are payable to (redeemable by) whoever happens to be the bearer. Such bearer claims are transferable without the bank's knowledge or involvement – Alice simply hands them over to Bob – and can change hands repeatedly before being redeemed. Today some versions of "smart card" payments, namely those like Mondex which allow transfer of balances directly from card to card without the bank's knowledge or involvement, amount to the reintroduction of banknotes in digital form.

Banknotes may have evolved from the practice of making payment by signing over a deposit receipt or cashier's check. When such payments are foreseen, depositors could ask for deposit receipts in round denominations for convenience, and in bearer form, to streamline and certify the payment. Payment was streamlined because signing over is no longer necessary. It was certified in the sense that the bearer note is a claim against the bank only, and not against any account that might have insufficient funds, nor against any subsequent endorser. No one who accepts a banknote – unlike a deposit receipt that is successively signed over – needs to worry that the goodness of the claim depends on the funds of those who have previously held it, or that he or she might be called upon to make good on it for those who hold it subsequently. A banker is happy to comply with requests to

issue such claims, as a way of increasing his circulation and profits. According to several accounts, this was the path by which goldsmith's deposit receipts historically evolved into banknotes (Usher 1943, Richards 1965).

The widespread use of banknotes historically preceded the widespread use of checking accounts (Bagehot 1873). For most British banks, note circulation exceeded deposits up to 1850. For banks in other countries, the date at which deposits began to exceed notes in circulation came even later. If banknotes evolved from deposit receipts, however, deposits on some scale must have preceded the use of banknotes.

Banknotes historically have paid no interest, even in competitive settings where deposits have, because there seems to be no easy way to pay interest on a bearer instrument whose convenience rests on its circulating at face value. In smart card payment systems thus far test-marketed, card balances similarly do not bear interest. The view that bank-issued bearer claims should be expected to bear interest in a competitive banking system will concern us in the last chapter of this book.

Regular Par Acceptance

Suppose a payments system has a common standard money, arrived at in the Mengerian way, with many banks issuing redeemable currency and deposit liabilities, but each bank refuses to accept any other bank's liabilities at par (face value). Bank-issued money then has limited marketability. This section argues that the profit motive, without legal compulsion, will move the banks toward par acceptance of one another's currency and deposits. An important side effect is the formation of an institution for interbank clearing and settlement of currency and deposit claims. The exposition refers explicitly to banknotes, both for convenience and for historical applicability to systems dominated by notes early on, but the argument applies just as well to par acceptance of deposit claims and digital currency.

An individual who has come into possession of a sum of notes issued by Bank *X*, and who wishes to deposit the sum into her account at Bank *Y*, has two options when Bank *Y* does not accept *X*-notes. She may bear the expense of taking the notes back to Bank *X* for redemption in outside money, or pay a fee to a note-changer (in the form of a discount on the *X*-notes and possibly a commission) who purchases the *X*-notes for outside money or *Y*-notes. Either option is naturally more expensive, the farther the noteholder is from a redemption site for the notes in question. Given these costs, *X*-notes are likely to circulate readily only in the vicinity of Bank *X* offices. Coin or other brands of notes will be preferred for transactions elsewhere.

In this situation, there are at least three logical scenarios whereby the

pursuit of profit leads toward widespread par acceptance; Selgin and White (1987, pp. 225–33) discuss these scenarios and offer historical illustrations:

1 Banks as note-changers
2 Note dueling
3 Mutual par-acceptance packs

Banks as note-changers: Banks can out-compete non-bank note-changers because they have the advantage of being able to issue their own notes (or deposit balances) to purchase other banks' notes. Where a non-issuing note-changer must hold costly till money on the asset side of its balance sheet, an issuing bank can hold interest-earning assets, giving it a profit from "float" for as long as the notes issued remain in circulation. (The Suffolk Bank of Boston in the 1830s succeeded famously at this business.) By swapping *Y*-notes for *X*-notes, Bank *Y* can maintain a larger stock of its own notes in circulation. Where the transactions and redemption costs of note changing are low enough to be covered by the float profit from additional circulation, competition will bring the issuing banks' note-changing fee down to zero. That is, competition will bring the banks to practice par acceptance. (Bank *Y* would never offer to buy *X*-notes at a price above par, because the note-changing customer could make arbitrage profits by turning around and immediately redeeming the *Y*-notes issued.) If all banks are thus drawn into zero-fee note-changing, mutual par acceptance develops de facto.

Note dueling: Bank *Y* may accept, or even aggressively purchase, *X*-notes, and then, suddenly, return a large quantity to Bank *X* for redemption in reserve money, hoping to gain a greater share of the banking market by embarrassing its rival. The trouble with this tactic is that two can play it. Bank *X* can collect and redeem *Y*-notes, both to return the damage, and to replenish its own reserves. Vigorous pursuit of the tactic on both sides (known historically as "note dueling," and practiced for example by the first and second chartered banks in Scotland upon the entry of the second) may drive note-changing commissions to zero. In a repeated game of this sort, however, a non-aggressive "tit for tat" strategy (return an opponent's blows in kind, but meet cooperation with cooperation) should evolve (Axelrod 1984). When both sides practice note dueling, both find it ineffectual and expensive. Greater non-earning reserves must be held at all times to meet a rival's large redemption demands that may arrive at any time. If neither party can win the duel, both should eventually recognize that a regular, and amicable, exchange of collected notes would benefit both by allowing them to economize on reserves. Fees will nonetheless remain zero as a defensive measure, allowing each to collect enough rival notes to safeguard its reserves.

Mutual par-acceptance pacts: As the fee for buying *Y*-notes with *X*-notes

falls, and *a fortiori* as it reaches zero (Bank *Y* offers commission-free par acceptance of *X*-notes), the cost of accepting *X*-notes falls and thus the circulation of *X*-notes grows relative to the circulation of outside money. This result is no part of the intention of Bank *Y* in the above scenarios, and may not be anticipated before the fact. (The Suffolk Bank was evidently surprised to find that by purchasing country bank notes at par, it was not vacuuming them from circulation but, in fact, encouraging their wider acceptance and use.) Bank *Y* aims only at an expanded circulation of its own notes, which indeed is also a result. Banks that accept other banks' notes at par improve the circulation both of their own notes, and of the notes they accept. If two banks both recognize *ex ante* the availability of these circulation gains from mutual par acceptance, they may explicitly enter a pact to accept one another's notes at par. Par-acceptance pacts among pairs of provincial Scottish banks provide historical examples of such agreements. Acceptance at par in a wider area increases the marketability of each brand of notes, and thereby the quantity willingly held by the public. The same logic explains the recent spread in the USA and elsewhere, of agreements among banks to form networks of mutual acceptance for cards giving access to automatic teller machine services. By participating in an ATM network, a bank improves the accessibility of its own deposits, and thereby attracts more depositors.

Par acceptance, developed through any of these routes, is generally more profitable, the wider its scope. The potential gains are not exhausted until all reputable banks practice par acceptance toward all others. Thus every bank's liabilities come to circulate at par throughout an economic region. The boundaries of the region will lie where the circulation-enhancing benefits of membership (presumably declining at the geographic margin as distance from the financial center increases) become equal to the transaction, transportation, and administrative costs of membership (presumably rising at the margin). As transaction and transportation costs secularly fall, the par-acceptance region expands. Par circulation of notes became nationwide with the spread of railroads in the nineteenth century. ATM networks are rapidly becoming global today.

Clearing Arrangements

Following any of these scenarios, Bank *Y* will be collecting *X*-notes, and Bank *X* will be collecting *Y*-notes, during the course of the business week. Each bank will want to redeem the collected notes, rather than to pay them back out again (as the whole profitability of the arrangement comes from placing and maintaining more of its *own* notes into circulation) or to accumulate them indefinitely (reserve money is more useful, and no more costly

to hold). A regular meeting for bilateral redemption, where *X*-notes are traded for *Y*-notes and the difference settled in reserve money, or some agreed substitute, will be arranged when the banks find that it is cheaper than unilateral or irregular redemption.

The practice of regular bilateral redemption may emerge without the banks' management planning it. If note-porters are sent from each bank to redeem at the other bank's counter, and they happen to meet, it should readily occur to them that an exchange of notes would save them time and the effort of lugging a great deal of gold (all but the difference) back home. They will arrange to meet regularly at a specified time and place to exchange notes. Bank management will endorse the arrangement not only because it saves transportation costs, but also because bilateral netting (using Bank *Y*'s claims against Bank *X* to offset Bank *X*'s claims against Bank *Y*) allows smaller reserves to be held.

The gains in going from unilateral to bilateral note-exchange are further extended in going to multi-lateral exchange in a system of more than a few issuers. Time and transportation costs are further economized by having one all-encompassing meeting rather than numerous pairwise meetings, and the holding of reserves can be further reduced with multilateral netting of claims that in pairwise clearing would have to be settled in reserve money. Multilateral exchange may evolve from bilateral exchange in the same way that bilateral exchange evolves from unilateral exchange. The note-exchange agent for Bank *X*, having concluded his or her regular exchange session at Bank *Y*, may happen to meet the agent for Bank *Z* arriving for his or her meeting at Bank *Y*. There are the economies just mentioned in combining the two meetings, and absorbing as well the regular pairwise meeting between *X* and *Z* agents. (The London note-exchange reportedly grew out of such note-porters' meetings in pubs.) Unified computation, and settlement of combined net clearing balances, can economically replace three bilateral exchanges.

Other banks may be invited to join the clearing sessions subsequently, either individually or through combination with a similar multi-sided clearing group. Eventually, all reputable banks within the par-acceptance region will be linked through a single clearinghouse, or through a small number of subregional clearinghouses that regularly clear against one another. The development of clearing arrangements in Edinburgh, London, and New York all conform to this general pattern. The final outcome – a unified clearing system encompassing all banks – is not part of any bank's initial design. Each aims only at increasing the market for its own liabilities, and at economizing on redemption and reserve-holding costs. Systemwide par acceptance, and its embodiment in the clearinghouse, in this sense, represent a spontaneous institutional order.

The simplest and initial way of settling interbank clearing balances is through the physical transfer of outside money at the end of the clearing session. Echoing

the original development of deposit-transfer banks to provide payments more cheaply between bank customers like Alice and Bob, the banks may find it economical to make interbank payments by means of a banker's bank. They can settle up by transferring claims to outside money held in the clearinghouse vault rather than by physically carting outside money around. Clearinghouse association (CHA) banks in US cities in the nineteenth century issued claims in the form of paper certificates to be used for interbank settlements. Other CHAs have used clearinghouse deposits for the same purpose.

Historically, CHAs have been known to take on functions additional to their core function of economically clearing and settling claims among banks. One is the sharing of information on loan defaulters, passers of bad checks, and the like. More significant are certain "hierarchical" functions associated with policing the soundness of member banks. Chapter 4 discusses the possible connection between such functions and the emergence of central banking.

The Path to Fiat Money

At this point, we can take stock of the spontaneously or "naturally" developed monetary system so far described. The definitive money is specie. Except in interbank settlements, transactors commonly make payments using bank-issued currency and transferable deposits. A specie unit is the unit of account. Bank-issued money is denominated in the specie unit, and is widely accepted at par. All banks are linked into a unified system by one or more clearinghouses. These outcomes are not purely theoretical, but could be seen historically in banking systems that were free of significant legal restrictions.[13]

Is there a spontaneous or market-driven path from this system to the non-commodity, or fiat, standards that prevails today? No. If any single bank in the system were unilaterally to stop redeeming, it would have breached its contracts with its customers. If it were to announce in advance that it would stop redeeming next month, holders of its notes and deposits would redeem them all before next month, and would take their business elsewhere. (Alternatively, if the bank tried to replace ordinary open-ended notes and deposits with new liabilities whose redeemability was scheduled to expire on a specified date, nobody would take the liabilities as the date approached.) The other banks and the public would reject the irredeemable liabilities because without redeemability at par for specie, there would be no assurance of continued par value in terms of the specie unit of account. (Chapter 12 considers the question of private irredeemable money in more detail.)

[13] See Dowd (1992b) for case studies.

The forces that lead to convergence on a common monetary standard, as in Menger's account, continue to operate once a standard is reached. Nobody wants to make trading harder by offering or accepting only a non-standard money, different from that routinely accepted and offered by others. Consequently, nobody would want to go first in switching to a completely novel monetary standard, even if he were persuaded that, in theory, it would work better supposing that *everyone* switched. (The policy implications of this phenomenon are discussed in chapter 5). If nobody goes first, the switch never occurs.

If all the banks, together, could coordinate a simultaneous switchover to a fiat standard (a very big "if," but banks did coordinate the beginnings and ends of temporary systemwide suspensions of payments during a few of the nineteenth-century US banking panics) the new standard might stick. However, it is not clear what market forces would compel banks to want to make such a move. Also, if it meant breaching pre-existing redemption contracts (as suspensions of payments historically did), it would not be a voluntary switch by the users of money.

In historical practice, a nation's switch to fiat money was typically made by the central government first granting a legal monopoly of note-issue to a single institution, a central bank, whose liabilities became as widely accepted as specie, and displaced specie as the reserves for other banks. The government then suspended, permanently, the redemption of the central bank's liabilities. With their permanent suspension, central bank notes and deposits became a fiat base money. The fiat-money unit correspondingly became the unit of account. Typically, the central bank for continuity's sake retained the old specie unit name (e.g. "dollar"), which was printed on the notes in circulation at the moment of suspension, while severing its specie definition. The now-irredeemable notes can continue to circulate because they are familiar, and the practice of continuing to accept them is self-reinforcing: it is not in any one trader's self-interest to refuse them if she expects others to continue accepting them.[14]

Thus, fiat money is possible where paper banknotes had previously gained

[14] In addition, the government can reinforce their continued acceptance by making the now-irredeemable central bank liabilities

1 *publicly receivable* – taking them for tax payments and for purchases from state enterprises;
2 *legal tender* for payment of old debts contracted in the unit of account;
3 *forced tender* in all domestic exchanges, including spot transactions that traders would rather conduct in another currency and repayment of old debts specifically denominated in metallic units.

As a final step, the government can, as the US government did, require the public to turn in its specie.

common acceptance as redeemable notes.[15] Likewise, to launch a new fiat money today (for example, in former Soviet republics), it must at first be made redeemable for the prevailing money (the ruble). Selgin (1994a) likens initial redeemability to a "launching vehicle" that can fall away once the new currency gets into orbit.

Spontaneous Separation between the Media of Redemption and Account?

A few authors, in recent years, have argued that the story of market monetary evolution should not end with banks operating on a silver or gold standard. Absent legal restrictions, Kevin Dowd (1996, pp. 14–18) argues that the next logical market-driven steps would be:

1 a discontinuation of direct redemption for precious metal, and
2 a switching over of the medium of account to a multi-commodity standard.[16]

Dowd's scenario for the discontinuation of direct redemption runs as follows. Once the stage is reached where gold coins no longer commonly circulate, because everyone prefers to use bank-issued money for all transactions, the public will no longer care about having the option to redeem for gold as such. Because a bank can pay its depositors a higher return by replacing all its non-interest-bearing gold reserves with interest-bearing assets, competition would compel banks to make the switch, and, correspondingly, to offer redemption in equivalent financial assets rather than gold itself (call this "indirect redemption"). At this juncture, a weight of gold would still define the unit of account, but gold would no longer serve as the medium of redemption. If, for example, one "dollar" is one-twentieth of a Troy ounce of gold, the holder of a $100 banknote could redeem it for financial assets (perhaps blue-chip bonds or equity shares) equivalent at market prices to five ounces of gold. Ounces-worth or "indirect" redeemability,

[15] Two historical cases are instructive here. When the Bank of England suspended payments from 1797 to 1819, Northern Ireland remained on a specie standard because banknotes did not yet commonly circulate there. For the same reason, California remained on a specie standard while the rest of the Union went on to an irredeemable "greenback" dollar standard during the American Civil War.

[16] Cowen and Kroszner (1994, pp. 38–44), using quite a different argument, imagine evolution toward a variety of parallel standards. Greenfield and Yeager (1983) propose a multi-commodity standard, but do not claim that spontaneous market forces are enough to ensure its adoption.

like direct redeemability, satisfies the customer's demand for contractual assurance of the value of bank-issued money. Though the typical retail customer might never exercise the new redemption option, just as he or she had come never to exercise the old, the option maintains an avenue for arbitrage. It keeps the market price of gold, quoted in any particular brand of dollar-denominated banknotes and deposits, tied to the definitional gold content of the dollar.

Is this scenario convincing? There is historical evidence that, in a gold-standard country with a sophisticated banking system, the use of gold coins as a hand-to-hand medium of exchange has indeed tended to diminish substantially. There is certainly evidence that banks take steps to economize on their holdings of non-interest-bearing reserves. Dowd (1989, p. 155) cites the case of the nineteenth-century Scottish banks whose customers were often (which is not the same as *always*) satisfied to redeem their liabilities for drafts on London correspondent banks (presumably because they had payments to make in London) rather than gold. Those Scottish banks could hold much (which is not the same as *all*) of their reserves in the form of interest-bearing accounts in London. The case suggests that, if customers and their banks sometimes mutually prefer redemption in something other than gold coins, then a competitive bank would offer its customers that option *in addition* to gold redemption.

There are, however, no known historical cases where competition led banks to reduce their gold reserves literally to *zero,* and to remove entirely the option of direct redemption in the medium of account from their banknote and deposit contracts. Thus, we have little reason to be confident that the public would, in fact, welcome the elimination of direct redemption. At the gold reserve ratios observed historically in the most sophisticated banking systems (2 percent and less), only *very* small increases in deposit yields are available by reducing reserves the rest of the way to zero. Unless the option of demanding gold rather than financial assets were completely valueless, it is doubtful that the public would welcome the final reduction of reserves to zero, and the complete elimination of the direct redemption option from bank liabilities.

The heaviest users of the redemption option are not retail bank customers, but the banks themselves: every day, banks redeem huge volumes of claims on one another at the clearinghouse. Whether direct redemption would spontaneously disappear thus depends on whether banks themselves would welcome the switch by their clearing partners, and would agree to accept settlement in financial assets (bonds or equity shares) rather than in the medium of account (e.g. gold or clearinghouse claims directly redeemable at par for gold). As noted above, banks historically found it convenient to settle net clearing balances by transferring claims to precious metal, kept in the clearinghouse vault, rather than by physically carting bullion, or bags of coins, to and fro. However, there are no known historical cases where member banks agreed to

reduce the clearinghouse's vault cash to zero, and to settle in clearinghouse claims that were not directly redeemable for outside money.[17] A straightforward explanation is that no historical clearinghouse in a gold-standard banking system could do without gold reserves because its members had to be prepared to pay gold to the rest of the world. No clearinghouse embraced all the banks in the world, and gold remained the medium of settlement between banking systems.[18] A move to indirect redemption is conceivable only for the clearinghouse of a closed economy, which in the world of an international gold standard means a global clearinghouse.

Members of a clearinghouse might agree to have it hold fractional reserves, so that their clearing account balances could bear interest. (The New York Clearinghouse in the nineteenth century, however, typically held 100 percent reserves.) This does not imply an end to direct redeemability, or settlement with financial assets. The convention of settling in one of the two traditional ways, either by transferring physical units of outside money, or by transferring directly redeemable (and hence par-valued) clearin house claims to outside money, reduces transaction costs because it transfers a single homogenous asset with an unambiguous unit-of-account value. All parties can agree that a $10,000 clearing balance is settled with $10,000 in full-bodied coins or a claim redeemable at the clearinghouse for $10,000 in coins. An imagined system of settling with financial assets, by contrast, appears to face banks with the problem of continually negotiating agreements about *which specific assets* are acceptable, and for each asset *at what price* within the interval bounded by the asset's current bid and ask prices (which prices would have to be continuous tracked in a system of continuous settlement). It does not appear that these agreements could simply be reached once for all time, because the set of financial assets available is continually changing. Even items within the set change in their characteristics. In the bond market, for example, new bonds are being issued, old bonds are being retired, extant bonds are shrinking in duration and, hence, in interest-rate risk as they approach maturity, the default

[17] Dowd (1989, p. 96) wishes to attribute the non-realization of his scenario to state intervention, but it is not clear that private clearinghouse associations were everywhere prohibited from taking such a step.

[18] Describing the international payments system in the early part of the twentieth century, Ludwig von Mises (1980, pp. 325–6) observed that although "the clearing system has without difficulty transcursed political boundaries and created for itself a world-embracing organization in the international bill and check system," there were still no bank-issued payment media "that are recognized internationally and consequently able to take the place of [metallic] money in international trade for settling the balances that remain over after the clearing process." The absence of global branch banking and globally accepted bank liabilities was, of course, at least partly due to legal restrictions.

risk associated with particular issuers is varying, and the liquidity of particular securities (as indicated by the size of their bid-ask spreads) is changing.[19]

Dowd's scenario for the spontaneous mutation of the medium of account runs as follows. So long as the unit of account is defined as a fixed weight of gold, the price level is subject to disturbance by supply and demand shocks in the market for gold. (This feature of a gold standard is analyzed in chapter 2). "A time would therefore come when the banks would decide to reduce price-level instability" by replacing gold with a medium of account "with a more stable relative price," most likely consisting of "a basket of goods and services" (Dowd 1996, p. 16).[20] The banks would arrange to act in unison. The public would go along because they prefer greater price level stability.

The question poses itself in this case as well: if this unit-of-account switch is a move worth making, why have banks and clearinghouses historically avoided it? One explanation, parallel to the previous discussion, is that no historical clearinghouse found it advantageous to switch its medium of account while the rest of the world stayed on gold. If no single clearinghouse (or its members' customers) would wish to "float" against the rest of the world, then a move to the new regime – from the status quo of an international gold standard – is again conceivable only for a global clearinghouse. Chapter 5 below discusses in more detail the idea that no one finds it worthwhile to switch the unit of account unless everyone switches together.

It is not necessary to switch the unit of account, however, for a bank or clearinghouse to insulate the value of its liabilities from changes in the relative price of gold: it could index its redemption rate. To allow continuous adjustment of the redemption rate, the indexation could be to the price of a basket of standardized commodities continuously traded on organized exchanges.[21] The question becomes, then, why did banks and clearinghouses on the gold standard avoid indexation of their liabilities? (White 1990, p.

[19] Why wouldn't it be enough for the banks simply to agree once-for-all to accept settlement only in *default-risk-free* securities, like present-day US Treasury bills? First, default risk is not the only relevant risk for banks holding bonds. Second, under a commodity standard, there simply are no default-risk-free bonds. Under a gold standard, for example, no government can print up gold to repay its gold-denominated bonds. In this respect, the euro standard acts like a commodity standard: no participating national government can issue default-risk-free bonds denominated in euros, because no nation can print euros to redeem its bonds. (Only the European Central Bank can print euros.)

[20] This switch in the medium of account could, in principle, be made whether the banks practiced direct or indirect redemption. With the banks practicing direct redemption in gold, it amounts to switching from an unindexed to an indexed redemption rate. With indirect redemption, the banks would already be indexing the redemption rate in financial assets to the gold price of the assets; now they would switch to indexing in the basket price.

[21] This sort of indexation is discussed at greater length in chapter 12, in connection with the reform proposal of Greenfield and Yeager (1983).

197) Perhaps the theory of indexation and the organization of commodity exchanges were simply not developed enough by the time the gold standard was ended by other means. Or, perhaps indexation would not have been worth the bother because the instability of the relative price of gold (when gold is money) is not much greater than that of any feasible index basket that allows for continuous tracking.

1 Once traders begin using indirect exchange, why do they tend to converge on a single good as the commonly accepted medium of exchange? Could they converge on two or three goods?
2 If all goods were equally salable, would money still emerge out of barter?
3 Why did commodity monies, rather than fiat monies, historically emerge out of barter economies?
4 Consider a system where gold coins are the commonly accepted medium of exchange. Absent legal restrictions, would private issuers have incentives enough to establish a *uniform* monetary standard? Or, would a variety of coins, of different weights and finesses, circulate?
5 Why have traders, historically, often preferred to use claims on banks (like banknotes and checks) even though the claims might be dishonored, rather than precious-metal coins?
6 Why has the unit of account typically been some quantity of the commonly accepted medium of exchange?
7 J. Huston McCulloch (1982, pp. 6–7) has argued that "the development of money is very similar to the development of language." There is a tendency, in any society, for *one* verbal utterance to be singled out as *the* spoken word for "fire," but it is "fundamentally arbitrary which grunt or series of noises" people adopt for that purpose. Likewise, there is a tendency for one commodity to be singled out as *the* common medium of exchange, but "which commodity is singled out is largely a historical accident." In what respects do you agree, in what respects do you disagree, and why? Cite theoretical, and historical, support for your position.
8 "As the supply of metals in the world has repeatedly been insufficient to meet the increasing need for money, the use of paper [currency] has spread." (Cribb 1986) Do you agree or disagree with Cribb's explanation for the historical shift from coins to banknotes? Explain why.
9 John Browning and Spencer Reiss (1998) define "feedback, positive" as "success that breeds success" and elaborate:

> Positive-feedback loops create a winner-take-all world: whoever or whatever starts ahead gets further and further ahead. . . . Incumbents literally own their markets. . . . Once a positive-feedback loop does kick in, the result can be awesome concentrations of economic power. Bill Gates once explained it this way, talking about Windows: "Momentum creates momentum. If you have volume, then people write apps. If people write apps, you have momentum." And if it goes on long enough, you have $40 billion.

(a) Is Menger's theory of the origin of commodity money a story about a positive-feedback loop? How or how not?

(b) Does convergence to a common commodity money standard lead to an incumbent owning the market, or to a concentration of economic power? How or how not?