Solving the problem of small change

by François Velde, senior economist

In the last hundred years, monetary systems have evolved from commodity-based, meaning their value is tied to a commodity like gold, to fiat-based, meaning their value is not tied to anything. Why did this happen? What was wrong with the old commodity-based systems? A recent book explores these historical developments and shows how the problem of small change played a crucial role in the introduction of fiat money.

Our experience of money in the twentieth century has been very different from the past. Figure 1 illustrates the point by showing the price level over a long period in four countries: England, France, Spain, and the United States. The base year for the index is 1913, so all series coincide in that year. The most obvious feature of this graph is the big change around 1914, which marks the end of the gold standard, a commodity-based monetary system. Since then, we have used fiat money, which is not tied to the value of a commodity. As figure 1 shows, the price level was roughly constant for three hundred years before 1914—with some fluctuations, to be sure, but nothing like what we saw after 1914.

In a radical departure from previous practice, the price level was in some sense set loose and allowed to rise. Did the idea of fiat money occur suddenly, or was the groundwork laid for it earlier?

A recent book by Sargent and Velde (Thomas Sargent and François Velde, *The Big Problem of Small Change*, Princeton University Press, 2002) explains a problem that bedeviled monetary systems for hundreds of years, the problem of small change. They show that: 1) we had been learning about fiat money for a long time before 1914; and 2) we learned as we tried to cope with the problem of small change. This *Chicago Fed Letter* provides an overview of some of the issues involved.

For hundreds of years, supplying small denominations of currency was difficult. This may seem like a trivial question to us today, because the problem has been solved; but in the past, this created many recurring problems, such as shortages of small denominations, variable exchange rates between coins of different sizes, bouts of inflation, invasions of foreign coins, and depreciation of the currency.

As societies struggled to solve the problem, they moved from a commodity money system, in which the value of money was tied to that of a commodity like gold or silver, to a fiat money system, in which money is essentially valueless (paper) but is valued in the marketplace. The solution, which was considered standard at the end of the nineteenth century, was to make the small denominations fiat but “convertible” into the (gold-based) large denomination.
Making smaller denominations “fict” meant that their intrinsic content was substantially lower than their face value, which is obviously the case today with paper money. But it also involved pegging the exchange rate between small and large coins. In the process, societies learned a lot about managing fiat money, its advantages, and its risks.

**What’s the problem?**

Why was commodity money so bad at producing small denominations? Consider the case of a single coin, for which the properties of a commodity money system are fairly transparent.

The main property of a commodity money system is to tie the price level to an external anchor, the value of the commodity on which it is based. It does so by putting limits on the value of the coin, which is inversely related to the price level. Specifically, the system gives a lower limit and an upper limit to the value of coins.

The lower limit comes from the coin’s content. If the market value of coins containing 1 oz of silver were to fall below the value of 1 oz of silver, it would be profitable to melt them down for their content; this would reduce their quantity and drive up their value.

The upper limit arises because of the way in which coins are produced. In the classic medieval system, which remained in place until the advent of fiat money, the mint (where coins were produced) stood ready to convert metal into coins at a set price. If the value of a coin were to rise above the price paid by the mint, it would be profitable to bring metal to the mint and turn it into coins.

These limits put constraints on the price level that are essentially tied to the value of the metal used in the coin. How far apart these bounds stand is determined by the seigniorage rate, the premium charged by the mint to turn metal into coins. If the mint has to break even (and it usually did), the seigniorage rate has to cover production costs, the costs of actually making the coins.

Between these limits, the value of coins (the price level) can be thought to fluctuate according to the typical “quantity-theory” relation. The reason that coins exist as coins rather than as metal is because they facilitate transactions. A given (real) volume of transactions necessitates a certain real value of coins. If the volume changes, and until the nominal, or physical stock of coins changes through minting or melting, the real value of each coin has to change. Thus, variations in the volume of transactions lead to proportionate variations in the value of coins. But large variations automatically induce the appropriate response by driving the value of coins to the lower limit (where melting occurs) or the upper limit (where minting occurs).

The quantity of money automatically adjusts to maintain the price level within its bounds, and there can never be a serious “shortage of money.”

At the starting point in European monetary history, in 800 AD, there was indeed only one coin in Europe, the penny. Later, larger coins were introduced. Two things happen, on the supply side and on the demand side. I use the words “pennies” and “dollars” to refer to small and large coins.

On the supply side, there are now two sets of constraints on the value of money, where money now takes the form of pennies (small denominations) and dollars (large denominations). They arise in just the same manner as in the single-coin case, as upper and lower limits on the value of each coin. But the limits may not be consistent with each other, because pennies are more expensive to make.

Thus, if the mints are required to break even (indeed, they were expected to yield a monopoly profit to the government), then they face the following dilemma: either charge more to make pennies than to make dollars, or charge the same but make pennies with less metal relative to dollars. In the first case, the risk is that, when money becomes dear, people will come to the mint to make dollars, but not pennies. If the government wants pennies, it has to make them on its account, and at a loss. The risk in the second case is that pennies are then “light,” that is, they contain proportionately less metal than they ought to compared with dollars. If the value of coins rises enough, then dollars will be melted first because they are relatively more valuable and only pennies will remain.

But the demand side has implications as well. To see this, suppose that small and large coins can be used for large purchases, but only small coins can be used for small purchases. This is a simple way of creating a demand for different denominations, and opens the possibility that the existing stock of pennies might suddenly become insufficient for the transactions that require them.
Why wouldn’t such a shortage cure itself, as in the one-coin system? Pennies will be produced only if their value rises enough. The value of money depends on the whole stock of coins, both pennies and silver, and there may be enough dollars around that the value of money (dollars or pennies) doesn’t rise enough. And there is the additional problem that the mint might be forced to charge more for pennies than for dollars, discouraging anyone from buying pennies.

This suggests that the relative price of pennies to dollars should adjust, so that pennies become more expensive relative to dollars, and their value rise to the minting point, even if the value of money (the inverse of the price level) doesn’t adjust. But in fact, as Sargent–Velde find, pennies depreciate during a shortage. This is a counter-intuitive result, but there is sound intuition for it.

The basic idea is to remember that money is an asset, and different monies will be compared by their rates of return. Most of the time, pennies and dollars are substitutes, and people don’t care which they hold. But in times of shortages, prices have to signal to people that they must economize on their holdings of pennies. For regular goods, this is done by driving up the price of the good in short supply. But for assets, this is done by making pennies less attractive than dollars as a store of value: that is, by making them depreciate.

Another way to see the same idea is to think of money as providing store-of-value services and liquidity services. All forms of money should provide the same total services, or else they wouldn’t all stay in circulation. During shortages, pennies provide more liquidity services relative to dollars, because they alone can be used to buy small goods, and there aren’t enough pennies to buy all the desired small goods. So pennies must provide less store-of-value services: i.e., depreciation and postponing the point where pennies are melted down. If pennies are debased enough, it can also make it profitable to mint pennies. The historical record shows repeated debasements of small denominations prompted by currency shortages over centuries.

The long-term cure, which was a standard formula in monetary textbooks around 1900, was to take all the small denominations out of the market-driven supply mechanism: make them token, so they are cheap to produce, and make them convertible on demand, so they don’t lose their value and so that the government always supplies the right quantity. This amounts to pegging an exchange rate. The gold standard of 1900 is just that, with silver dollars and subsidiary coins being overvalued with respect to their content, and the Treasury legally obliged to maintain their parity with the gold coinage.

Simple as it may seem, this solution had three main prerequisites, which may explain why it took so long for the solution to be implemented. The first prerequisite is a technological one. Making coins token creates a big incentive for counterfeiters, so the government needs to either devote resources to enforcing the law or possess a technology that makes its coins harder to imitate.

The historical record shows repeated debasements of small denominations prompted by currency shortages.
for money balances, and real balances of these coins hit a ceiling. The nominal balances kept going up, as Venice continued to print them, and inflation ensued. Authorities in Venice stopped issuing torneselli, and then after a while devalued them to their intrinsic content, which is the downward shift in nominal balances we see at the end. Similar experiments with overvalued copper coinage were carried out, on a vastly larger scale, in seventeenth-century Spain and Germany, resulting in spectacular bouts of inflation.

Conclusion
Sargent–Velde show many examples of how governments have been experimenting with fiat-like money for centuries. Not all experiments resulted in inflation, and there are early examples of attempts at implementing the solution to the problem of small change, although counterfeiting repeatedly stymied them. And, as figure 1 suggests, learning about how to manage fiat money continued in the twentieth century. But enough elements of modern fiat money had been put in place by 1914 for the likes of Irving Fisher to contemplate the gold standard and propose that the U.S. currency’s last link to a commodity be removed.