Central Bank Digital Currencies: An Old Tale with a New Chapter

by

Michael D. Bordo, Rutgers University
William Roberds, Federal Reserve Bank of Atlanta

Griswold Center for Economic Policy Studies
Working Paper No. 323, May 2024
Central bank digital currencies: an old tale with a new chapter*

Michael D. Bordo
Rutgers University, NBER and the Hoover Institution, Stanford University

William Roberds
Federal Reserve Bank of Atlanta

This version: April 30, 2024

Abstract: We consider the debut of a new monetary instrument, central bank digital currencies (CBDCs). Drawing on examples from monetary history, we argue that a successful monetary transformation must combine microeconomic efficiency with macroeconomic credibility. A paradoxical feature of these transformations is that success in the micro dimension can encourage macro failure. Overcoming this paradox may require politically uncomfortable compromises. We propose that such compromises will be necessary for the success of CBDCs.

JEL codes: E42, E58, N10

Keywords: monetary systems, banknotes, central banks, digital currencies

* An earlier version of this paper was prepared for a conference held in honor of Angela Redish at the University of British Columbia, October 14, 2022. The authors are grateful to conference participants and to an anonymous referee for helpful suggestions and corrections. The opinions expressed in this paper are those of the authors and not of the Federal Reserve Bank of Atlanta or of the Federal Reserve System.
Every year a new cohort of economists begins their professional careers with a trip into a fantasy land, one more unlikely than anything Walt Disney ever imagined. In this make-believe world, known as the Arrow-Debreu model, any good may be effortlessly exchanged to obtain another: socks are swapped for automobiles, peanut butter for electricity, labor hours today for medical services ten years from now. Yet despite its evident lack of verisimilitude, this abstract “theory of value” underpins much of modern economics and finance.

The idealized nature of the Arrow-Debreu model also offers insight into the function and evolution of money. A complication that makes the model seem unreal is the phenomenon of adverse selection, famously described by George Akerlof in the context of used cars (“the market for lemons”; Akerlof 1970) and by Groucho Marx for new houses (“you can even get stucco”; Paramount Pictures 1929). Money, a pointless entity in the land of Arrow-Debreu, offers a popular, if partial solution to this problem by concentrating adverse selection on one side of a transaction: the buyer gets a good, the seller gets money. The value of the money received can be determined by “tale,” i.e., by simple enumeration. Acceptance of money is not so much driven by illusion as by informational parsimony.

An ideal money from this perspective would need to be a safe asset, one whose value is so obviously invariant that no questions are asked about its quality—an NQA asset in the terminology of Gorton (2017). The evolution of money is marked by a progression of technologies that attempt to better approximate this ideal, as was explained by David Ricardo in 1816:

All writers on the subject of money have agreed that uniformity in the value of the circulating medium is an object to greatly be desired. Every improvement therefore which can promote an approximation to that object, by diminishing the causes of variation, should be adopted. (Ricardo 1816, 7).

More succinctly, Ricardo noted (ibid.) that “in the medium of circulation, [the] cause of uniformity is [the] cause of goodness.”

Modern forms of money appear to approach Ricardo’s dream of perfect uniformity; barter exchange is rarely prohibited but people prefer to swap goods for money. The perceived safety of
monetary assets contributes to this preference but is clearly not sufficient; to function as money, a safe asset must have a second property of \textit{liquidity}—ready exchangeability for other goods once it is received as payment. Gaining liquidity can be challenging, as is evidenced by the repeated failure of the U.S. Treasury to introduce a very safe asset—the U.S. $1 coin—into general circulation. In more technical terms, new moneys must overcome the hurdles posed by “adoption externalities” and “usage externalities.”\footnote{These concepts adoption externality and usage externality are often employed in payment economics to describe chicken-and-egg problems faced by new forms of payment (see Hunt 2003). A technical debate exists as to whether these phenomena are bona fide externalities or merely network effects.} Established forms of money have found ways past these hurdles, so effectively that economists often ignore the frictions that these moneys address, and instead treat the world as if it were Arrow-Debreu.

However successful both in practice and in theory, money continues to derive its value from the underlying informational asymmetry of real-world transactions. Trust therefore remains key to an asset’s monetary function, meaning that all money is to some extent fiduciary. This fact gives rise to a phenomenon we label “the monetary paradox”: the closer an asset comes to an ideal money (perfectly safe and liquid), the fewer questions get asked and the more opaque it becomes; yet this same opacity increases the profits to money creation and thereby, the temptation to undermine the safety on which money-ness depends. In the words of Milton Friedman:

So long as the fiduciary currency has a market value greater than its cost of production—which under favorable conditions can be compressed close to the cost of paper on which it is printed—all individual issuers have an incentive to issue additional amounts. A fiduciary currency would thus probably tend through increased issue to degenerate into a commodity standard—there being no stable equilibrium price short of that at which the money value of currency is no greater than that of the paper it contains. And in view of the negligible cost of adding zeros, it is not clear that there is any finite price level for which this is the case. (Friedman, 1960, p. 7)

In this essay, we will describe monetary history as a sequence of transformations, encompassing both virtuous improvements as envisioned by Ricardo and varieties of sin as described by Friedman. The result has been repeated rediscovery of the monetary paradox, with the consequence that progress has been nonlinear: every potential improvement has encountered a series of setbacks. We argue that successful monetary transformations have been ones where a
new monetary instrument, or a new way of managing existing instruments, can combine microeconomic efficiency gains with macroeconomic policies that do not undermine micro efficiency. Successful moneys are the product of such transformations. We then draw on this history to explore an emerging transformation, marked by the arrival of central bank digital currencies (CBDCs). We argue that CBDCs can best be understood as new chapter in an old tale, and that the technical novelty of this emerging form of money will not provide an easy escape from old patterns.

**Coins as the original transaction technology**

Theoretical models of payments begin by dismissing Arrow-Debreu. These models instead assume that transactional frictions, including adverse selection, are so extreme that no exchange can occur without some special arrangement (Kahn and Roberds 2009). This stark theoretical approach is supported by historical experience. With the notable exception of the Incas (Mann 2011, 84), early civilizations required some form of money to facilitate exchange. Surviving records of ancient Egypt and Mesopotamia indicate that these societies made use of state-sanctioned, book-entry payments (Davies 2005, Hudson 2020). Payment by transfer of monetary objects, however, is less labor-intensive than bookkeeping, and can sometimes be informationally equivalent (Taub 1994). Ancient societies across Eurasia sought to gain payments efficiency by trading in proto-monetary objects—cowrie shells, spades, hoes, axes, lumps of various metals (bronze, electrum)—before gradually adopting state-sanctioned, imprinted metal disks (i.e., coins) as money, a transformation that began as early as 750 B.C.E. (Davies 2005, Bresson 2021).

Monetary progress was never monotone, however, and the success of coined money within the ancient Old World was followed by setbacks during the Middle Ages. China, lacking access to large sources of precious metal, sought to augment its bronze coinage with state-issued paper money, first during the Southern Song dynasty (twelfth-thirteenth centuries; see Glahn 2016, 263-264) and again during the early Ming dynasty (fourteenth-fifteenth centuries; Glahn 2016, 287). Although China was at this time large enough and centralized enough to aspire to
paper-money issue, its rulers lacked a credible framework for managing these issues. Both of China’s early paper-money experiments ended in uncontrolled inflation, general chaos, and ultimately, demonetization of the paper involved.

In Europe, monetary exchange contracted with the collapse of the Roman Empire and disappeared in many places. From the seventh century, however, again money came into common use, at first in the form of “silver pennies”—coins of intermediate value (Spufford 1988). By the late Middle Ages—the starting point of this essay—monetary exchange was widespread, if not ubiquitous, within Europe and employed a variety of coins: “black money” (silver money of weak alloy) for small local transactions, “white money” (higher-alloy silver coins) for larger transactions between merchants, and beginning in the fourteenth century, gold coins for the largest categories of transactions (ibid.).

Modern economics refers to each of these types of coin as “commodity money,” a name that is sometimes taken to mean that these coins were just standardized chunks of bullion. Works such as Redish (1988, 1993, 2000) and Glassman and Redish (1988) illuminate the true nature of medieval and early modern coinage, which was decidedly more information intensive. Though the value of coins ultimately derived from their fine content, it was not practical to weigh and assay coins for every transaction. A definitive assay required melting down the coin. Successful coins were thus trusted, largely opaque instruments that traded above their putative metallic value and sometimes well above their legally assigned value in terms of the local unit of account. In modern terms, each coin had a “brand.”

The trust accorded these coins was not provided for free. Mints routinely extracted a form of profit known as seigniorage, which arose from the difference between the price the mint paid for precious metal deliveries and the nominal value of the coins it produced (Sargent and Velde 2002). The collection of seigniorage was seen as a privilege of the sovereign and as such, was often the most reliable revenue stream available to fiscally underdeveloped states. Collecting seigniorage was however a process with great potential for abuse.

Redish (2000, 61-62) catalogs the four most common ways for coin producers to cheat their customers. These were, in order of increasing subtlety: reducing a coin’s weight, reducing
its fine content, assigning a higher legal value to ("crying up") a coin, and introducing a new coin that resembled an existing coin but with a higher nominal value per unit of fine content. Each of these manipulative techniques were de facto debasements (reductions of fine content for a given nominal value), which profited both the coin producer (often, a private contractor) and the sponsoring sovereign. Glassman and Redish emphasize that these manipulations could also be benign, defensive responses to external forces such as clipping and counterfeiting, movements in the market gold-silver price ratio, or debasements of competing coins from neighboring jurisdictions.

In addition, rulers were aware that modest debasements could impart monetary stimulus to their local economies. A fourteenth-century French monarch, Philipp III of Navarre, received some advice to this effect from his advisor, Guillaume de Soterel. De Soterel observed that many merchants actually favored debasement, its fiscal burden notwithstanding, so long as it was practiced in moderation:

[This] sort of men are those who engage in commerce, who wish for another sort of money. That is a middle sort of money… Trade is always poor, except when money is in a middle state. (Leroy 1972, 110, cited at Spufford 1988, 305)

De Soterel recognized that this policy choice would not please everyone. Peasants and workers who had to discharge their obligations via physical labor would always prefer money of weak alloy (more rapid debasement); nobles living off fixed rents would always prefer strong money (no debasement). In a prescient formulation of Milton Friedman’s concerns, De Soterel warned that his “middle sort of money” should not be confused with grossly debased war coins, whose poor quality could undermine their monetary function:

[This] sort of money is desired by lords when they are at war, and he can thus strike coins as feeble as he likes to have the means to pay his troops to defend him and his people and his land. But at the end of the war, he ought to take this money in again. (Ibid., 306)

The most surprising fact about debasements, offensive and defensive, may be how often people ignored them. The informational advantage of a branded, opaque transactions instrument

---

3 Although prices did tend to adjust quickly in response to large wartime debasements; see Rolnick, Velde, and Weber (1996). Why people chose to readily accept debased, clipped, or otherwise flawed coins has been a subject of theoretical investigations, including Velde, Weber, and Wright (1999), Bignon and Dutu (2017), and Bajaj (2019).
was enough for people to tolerate, or as De Soterel suggested, sometimes even prefer a certain 
pace of endemic monetary degradation. This advantage was sufficiently large that an important 
advance in coinage technology, edge marking, went unimplemented for about 150 years (Redish 
2000, 54-61). Edge marking, also known as milling, enhanced monetary uniformity by reducing 
coin clipping. Sovereigns, mint masters, and market insiders (money changers), all of whom 
benefited from the variable quality of unmilled coinage, delayed the common use of this 
technology until the late seventeenth century. By that time, a more far-reaching monetary 
transformation was well under way.

**Transformation 1: from coins to paper**

In 1545, a man stumbled across a silver lode near Potosí, in modern-day Bolivia, 
ostensibly while looking for a lost llama (Lane 2019, 20-22). The chance discovery of this silver 
lode, the world’s largest, soon trebled global silver production (Soetbeer 1876, 8). Potosí was 
followed by other spectacular finds of silver and gold ore in the New World (TePaske and Brown 
2010, 56, 113). By the close of the eighteenth century, 70 percent of the world’s silver and 40 
percent of its gold had been supplied from American mines. This flood of precious metal is 
sometimes credited with causing a worldwide inflation (Hamilton 1934), but its more lasting 
effect was to turbocharge long-distance trade. Increased trade, in turn, made precious metal 
globally scarce and ironically, helped foster the development of European paper money.

The most important paper instrument for long-distance trade was the bill of exchange. 
Such bills consisted of an order by a merchant in one city to a merchant in another city, 
instructing the second merchant to pay out a certain sum to a named beneficiary by a fixed date.

---

4 Before the discovery of the Potosí lode, the value of world gold production slightly exceeded that of silver 
production (Soetbeer 1876, 7). Post-Potosí, however, silver dominated world precious metal production until the 
discovery of the California gold lodes in 1848. Silver’s relative abundance meant that most countries were on a 
silver standard, de jure or de facto, during the early modern period.

5 See Palma and Silva (2021). “Turbocharge” is a relative term. Early modern world trade ratios (annual exports plus 
imports, as a percentage of world GDP) remained in the single digits (Estevadeordal, Frantz, and Taylor 2003). By the 
turn of the twentieth century, this ratio grew to about 20 percent, and today it runs about 50-60 percent. The 
secular increase in trade has been an important driver of the monetary transformations described in this paper.
Bills of exchange allowed merchants to move large sums between cities without the physical transfer of precious metal. The value of bill payments was also high enough to justify the attendant bookkeeping costs (e.g., a bill was typically issued in triplicate). Credit extended via bills (which were typically payable with a lag) also had the advantage of exemption from the Church ban on interest. Bills thus expanded the reach of coins over time and space.

Once a bill of exchange landed in its destination city, however, there then arose the question of how the bill would be settled. In principle, settlements could be made in coin. In practice, however, merchants often preferred to settle in local paper money (Jobst and Nogues-Marco 2013). Forms of settlement included “fair money,” a practice of netting out obligations during trade fairs, using a common, essentially fictional unit of account (Börner and Hatfield 2017), and settlement on the ledgers of private banks, known as the “goldsmiths” in London (Quinn 1997) or “cashiers” in the Low Countries (Aerts 2011). Settling debts on paper further economized the use of precious metal and sidestepped the quality problems endemic to early modern coins.

From the fifteenth century onward, public banks assumed a larger role in the provision of local paper money. The first such institution was Barcelona’s Taula de Canvi, founded in 1401. The idea of public banking then spread to Mediterranean commercial cities such as Genoa and Venice, and in the seventeenth century, to northern European cities such as Amsterdam, Hamburg, and Stockholm (Roberds and Velde 2016a). By one count, there were 25 public banks operating in Europe at the close of the seventeenth century (Clapham 1945a, 3). Almost all these early public banks were municipally chartered, ledger-money banks. To pay with bank money, payor and payee (or their servants) literally walked to the bank and initiated the transfer of funds from the payor’s account to the payee’s. Every transaction was recorded on the bank’s ledger and

---

6 Bills of exchange could also circulate prior to settlement, often via endorsement, to the extent this was allowed by local legal systems. Such “negotiable” bills of exchange thus served as money; examples of this phenomenon can be found in places such as the seventeenth-century Low Countries (Heyvaert 1975), eighteenth-century England (Gorton 2020), and nineteenth-century Spain (Cuadras-Morató and Róses 1998).

7 Both the London goldsmiths and the Low Countries cashiers also issued circulating notes. These would not attain the wide circulation and broad acceptance of later issues by public banks.

8 The term early public banks includes banks owned by municipalities (as in Amsterdam), banks owned privately but operating under an exclusive charter (as in London), and at least one case (Naples) of chartered banking institutions operated by religious charities.
was witnessed by at least three sets of eyes, minimizing the potential for fraud. Safety of bank money was ensured, at least in principle, by convertibility of account balances to coin.

A challenge faced by every ledger-money type of public bank was how to ensure sufficient liquidity. To this end, public banks were granted different types of privileges, and most public banks tried to compel local merchants to execute large transactions exclusively in bank money, a strategy that was difficult to apply in practice. This type of public bank was most successful in Amsterdam (Dehing 2012, Quinn and Roberds 2014, 2019). There, ledger money became so well established that from 1683, the Bank of Amsterdam could limit convertibility of accounts and operate in fiat money. This money was used to support favored parties: the city government and the Dutch East India Company. Consistent with Friedman’s prediction, excessive lending to the latter led to the bank’s collapse in 1795 (Quinn and Roberds 2016).

Transformation 2: from ledger money to banknotes

Ledger-money (“giro”) transactions enjoyed some success, especially for large-value transactions. But giro transactions were also costly, inconvenient, and by their nature never fully private. These disadvantages led people in early modern Europe to experiment with circulating banknotes as an alternative form of paper money. The first, tentative use of hand-to-hand, public-bank instruments emerged in the late sixteenth century. By the close of the nineteenth century, banknotes circulated in all major European countries and had assumed a modern form: standardized, fixed-denomination bearer instruments issued exclusively by a central bank or by a government quasi-monopoly (Redish 1993, 82).

The intervening 300 years were not characterized by a smooth progression toward Ricardo’s uniformity, however, but by a sequence of monetary experiments, many of these ending badly. The failures of paper money in early modern Europe would not have surprised an observer familiar with earlier Chinese experiences. More surprising would have been the unlikely success of a paper-money regime in one country (Britain), which would emerge as a prototype for such regimes throughout the globe.
Some of the earliest surviving public banknotes are from Naples (Costabile and Nappi 2018). These notes resemble portable ledgers, with the name of the party to whom the note was issued recorded on the note and the names of all subsequent transferees, as well as the dates and amounts of transactions. More anonymous, bearer-type notes were issued by public banks in Amsterdam in 1660 (Dehing 2012, 92) and Stockholm in 1666 (Wetterberg 2009, 44–45). Bearer notes represented an information savings over earlier ledger-type notes, as only the obligation of the issuer was recorded on the note. These also had the advantage that they could be used privately, but this privacy proved difficult to manage. The Amsterdam and Stockholm note issues were plagued by insider fraud and failed the safe-asset test. These early experiments ended abruptly with the execution and jailing, respectively, of the culpable bank officials and withdrawal of the attempted notes.

Banknotes attained new respectability with the founding and rise of the Bank of England (“the Bank”). Drawing on the example of Genoa’s Casa di San Giorgio, the Bank was chartered in 1694 as a debt management agency. The privately-owned Bank sought to leverage its stockholders’ capital by the issue of various kinds of short-term bills and notes. These were intended to fund the Bank’s large holdings of state debt. The target market for the Bank’s short-term instruments consisted of London merchants who were used to dealing in financial claims, including earlier types of notes issued by private parties (i.e., goldsmiths). Public skepticism of this plan was reflected in the fact that the Bank’s initial charter ran for only 11 years. Skepticism also meant the Bank’s notes were not recognized as legal money, only claims to money (to coin held at the Bank); that is, the safety of notes was to be ensured by convertibility.

Following a somewhat shaky start, the Bank enjoyed legendary success. The Bank’s first types of notes—customized, handwritten, and variable-maturity—gave way to more convenient, preprinted bearer notes in fixed denominations, which were easy to transfer and were redeemable

---

9 These notes were however still banknotes in the sense that they were at all times obligations of the issuing bank and not the note holder.

10 In modern payments jargon, these early notes were token-based rather than account-based payment instruments; see, e.g., Gorton and Zhang (2022).

11 Richards (1934, 219-30). The Bank’s founding has sometimes been described as a debt-for-equity swap.
on demand. Note denominations were kept large, £20 being the minimum until 1759, this amount being equal to about 40 weeks’ pay for an ordinary laborer in 1700. Use of the notes was not compulsory, but from 1708 the Bank gained the privilege of being the only large note issuer in the London market, thereby helping the Bank establish a sufficient network (Richards 1934, 212). The notes did prove popular with London merchants and the Bank experienced little trouble gaining charter renewals from Parliament. The quid pro quo for each renewal was that the Bank take on more state debt, but by a combination of shrewd management and good fortune, these accommodations did not founder the Bank. By 1788, the Bank’s note-heavy balance sheet had swelled to seven times that of its most prestigious rival, the ledger-money Bank of Amsterdam (Roberds and Velde 2016b, 485). To admiring contemporaries, the Bank seemed to have hit on a magical recipe for a banknote money (Alexander Hamilton: “a vast fabric of paper credit … raised on a visionary basis”) that was safe, liquid, and immensely profitable, both to its issuer and to its sponsoring government (Adam Smith: [the Bank] “acts as a great engine of state”).

In reality, the Bank’s triumph depended more on moderation than magic. Its notes were simple IOUs but were seen as safe and liquid by the Bank’s wealthy, London-based clientele. Adherence to strict accounting standards largely eliminated insider fraud. Counterfeiting and other forms of fraud were also minimal in the early decades of the Bank (McGowan 2007). The seeming ease with which the Bank’s notes enhanced Britain’s fiscal capacity generated envy and eventually, imitation in other European countries. These countries discovered, however, that the real key to the Bank’s success was not the superficial uniformity of its paper money, but its unwillingness to take on more state debt than its note circulation could sustain. The Bank’s credibility was also enhanced by Parliament’s policy of “funding” (allocating specific tax revenues to) most of the debt held by the Bank (Broz and Grossman 2004).

12 See the Bank of England website <bankofengland.co.uk/about/history>.
13 Historical data on English wages are from the dataset “A Millennium of Macroeconomic Data,” which can be downloaded from the Bank of England’s website at <bankofengland.co.uk/statistics/research-datasets>
14 Renewals were granted in 1697, 1708, 1713, 1742, 1764, and 1782; see Broz and Grossman (2004, 51).
15 See Hamilton (1850, 164) and the Wealth of Nations II.ii.85.
England’s success story proved difficult to replicate. Sweden’s second attempt at a public bank, the *Riksens Ständer Bank* ("Bank of the Estates of the Realm," later renamed the Sveriges Riksbank) was chartered in 1668 and began (reluctantly) issuing large-value customized, notes in 1701 (Wetterberg 2009, 58). These were gradually shifted to low-denomination, fixed-denomination bearer notes, which enjoyed considerable success, especially as an alternative to the cumbersome copper coinage then prevalent in Sweden. In 1739, however, a new political party (the “Hats”) came to power in the Swedish parliament. The Hats were determined to exploit the Riksbank in order to support Sweden’s expensive military campaigns (Bordo and Levy 2021) and to offer mortgage loans to the nobility (Heckscher 1934). These activities were funded by the issue of additional banknotes, which became irredeemable from 1745 (Fregert 2014, 342). The result was a collapse in the foreign exchange value of the Swedish currency and ultimately (in 1765), a political backlash when a new political party (the “Caps”) came to power. The Caps contracted Riksbank lending, bought up many of its notes, reestablished redeemability, and collapsed the Swedish price level by 37 percent within four years (1765 to 1769), thereby plunging the country into deep recession (Edvinsson and Söderstrom 2010).16

The Riksbank survived its troubles and exists today as the world’s oldest central bank. Other early note-issuing public banks were not as fortunate. France’s *Banque générale* was founded in 1716 by a Scottish financier, John Law (Velde 2003). Law’s bank expanded quickly and attempted to fund the entire burden of French public debt via banknotes and an equity issue. When the equity shares did not sell, Law printed more notes and used these to pump up the shares’ price. Collapse ensued in 1720. Austria’s public bank, the *Wiener Stadtbanco* (Vienna Municipal Bank) was founded in 1706 (Jobst and Kernbauer 2016). To allay public distrust, this bank was chartered as a municipal (not crown) agency and funded itself not with banknotes, but with interest-bearing time deposits. Unrelenting fiscal pressure, however, resulted in the de facto nationalization of the bank in 1759 and the first issue of banknotes in 1762, as a measure to help finance the Seven Years War (1756-1763). This initial emission of banknotes was withdrawn and publicly incinerated four years later, but Napoleonic-era fiscal pressures led to more note issues,

---

16 Sweden witnessed a second round of disruptive note issue during the Napoleonic period (Engdahl and Ögren 2008). These notes were not issued by the Riksbank but by an agency of Parliament, the National Debt Office. The Debt Office notes were irredeemable but could be used to pay taxes, and these soon established themselves as a parallel currency.
suspension of convertibility, and ultimately, a paper-money quasi-hyperinflation. The Municipal Bank was declared insolvent in 1816 and its note holders endured a 92 percent write-down.

Direct emissions of notes by governments fared no better than those of most early public banks. In Britain’s North American colonies, a lack of natural deposits of precious metal and strong economic growth led colonial governments to issue various forms of paper money. These notes enjoyed some local liquidity (sometimes due to tax privileges) but most were not redeemable on demand and did not approach NQA status (Sylla 2006); one historian (Grubb 2018) has argued that prices of notes in one colony (Virginia) were so variable that they functioned more as zero-coupon bonds than money.

At the national level, highly inflationary note issues by revolutionary governments, Continentals in the United States and assignats in France, ultimately proved (nearly) worthless and did little to enhance the reputation of paper currency (Capie 1986, White 2009).17 A rare success story occurred in Scotland, where private banks issued stable-value, liquid notes during the eighteenth and early nineteenth centuries. The literature has debated, however, whether the Scottish success was due to astute regulation (White 1984) or oligopolistic industry structure (Goodhart 1988).

Chastened by earlier experiences, the framers of the 1789 United States Constitution barred U.S. states from any form of note issue. The first U.S. Treasury Secretary, Alexander Hamilton, tried to create a national paper currency by concentrating note issue in a privileged private institution, the (First) Bank of the United States (Cowen 2000). Chartered by Congress in 1791 and closely modeled on the Bank of England, the privately capitalized First Bank was able to avoid many of the mistakes of its cross-Atlantic cousins. It did not suffer insider fraud, gross mismanagement, widespread counterfeiting, or fiscal depredation. Its large-value notes were to some extent seen as safe assets, and, thanks to a national branch network, these achieved a wide circulation. The undoing of the Bank was its technical success. The jealousy of smaller, state-

---

17 Grubb (2023) asserts that the first emissions of Continentals were so hopelessly un-money-like that people treated them more as zero-coupon bonds than banknotes, as had occurred with earlier colonial issues. Continentals were later assigned legal-tender status as wartime fiscal demands persisted.
chartered banks and popular resentment of its foreign shareholders (many of these British citizens) led to the dissolution of the Bank in 1811.

Even the poster child of banknote issue, the Bank of England, was confronted with severe challenges during the Napoleonic period. Wartime fiscal demands led the Bank to suspend convertibility in 1797 and issue emergency money in the form of £1 and £2 notes, pushing the Bank into the unfamiliar business of retail payments. Organized criminals then unleashed waves of counterfeit small-denomination notes (Palk 2006, McGowan 2007). The need to obtain jury convictions of counterfeiters meant that the Bank was now confronted with a costly and difficult enforcement problem. New legislation passed in 1801 allowed the Bank to prosecute anyone caught with a counterfeit note, for any reason, with the more lenient (i.e., non-capital) punishment of deportation to Australia for 14 years upon conviction (“transportation”). Armed with this new law, the Bank took legal action against thousands of counterfeiters, many of them women, and was able to transport or execute the majority of them. The ferocity of this anti-counterfeiting campaign undermined the Bank’s standing with the public in general and with juries in particular. Convertibility of Bank of England notes was resumed in 1821 but the troublesome small-denomination notes had to be withdrawn, a capitulation to the forces of adverse selection.

Table 1 summarizes the note-issuing experiences of selected early public banks into a “report card.” The record of these institutions was hardly one of consistent success. Only in two locations (London and Naples) did public banknotes obtain a consistent passing grade.

---

18 At this time, £1 was equal to about two weeks’ wages for an average worker.

19 More specifically, the Bank had to pay rewards to police for turning over counterfeiters, hire its own network of prosecutors, select winnable cases for prosecution, and pursue these cases to conviction. Public sympathy for the typically underprivileged counterfeiters, who faced the death penalty if convicted, often resulted in jury nullification of the Bank’s efforts.
Table 1: Report card for note issues by early public banks

<table>
<thead>
<tr>
<th>Location</th>
<th>First note issue</th>
<th>Result</th>
<th>“Grade”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naples</td>
<td>1587</td>
<td>Early success with customized, ledger-like notes</td>
<td>B</td>
</tr>
<tr>
<td>Amsterdam</td>
<td>1660</td>
<td>Gave up note issue in 1673 due to insider fraud</td>
<td>D</td>
</tr>
<tr>
<td>Stockholm I</td>
<td>1661</td>
<td>More insider fraud, bank closed in 1667</td>
<td>F</td>
</tr>
<tr>
<td>London</td>
<td>1694</td>
<td>Success with large denominations; issues of small denomination notes failed due to counterfeiting</td>
<td>B</td>
</tr>
<tr>
<td>Stockholm II</td>
<td>1701</td>
<td>Initial success; later inflated a real estate finance bubble; unwinding crashed Swedish economy</td>
<td>C</td>
</tr>
<tr>
<td>Paris</td>
<td>1716</td>
<td>Attempted to finance entire French state debt; collapsed in 1720</td>
<td>F</td>
</tr>
<tr>
<td>Vienna</td>
<td>1762</td>
<td>Ran quasi-hyperinflation during Napoleonic wars; liquidated in 1816</td>
<td>D</td>
</tr>
<tr>
<td>Philadelphia</td>
<td>1791</td>
<td>Technical success provoked political jealousy; bank liquidated in 1811</td>
<td>D</td>
</tr>
</tbody>
</table>

Note: Grades assigned in Table 1 are authors’ subjective assessments based on narratives given in the text. Grades follow the U.S. system in which A is the highest and grades below C indicate failure.

Transformation 3: The Classical Gold Standard

The monetary situation in early nineteenth-century Europe and North America can be described as “orderly chaos” (Bordo and Redish 2016, 597). After the disruptions of the Napoleonic wars, many countries returned to the coinage standards of the early modern period. By historical happenstance, Great Britain was on a gold standard. France, and all of continental Europe with the exception of Portugal, as well as the United States, were on bimetallic standards.

---

20 As noted earlier, most early modern polities were on a de facto silver standard because silver was in greater abundance than gold. Great Britain was an exception because its coinage statutes overvalued its gold coin (sovereign) relative to silver, as compared to continental coinage statutes that set this ratio closer to prevailing...
A major contributor to this chaotic atmosphere was disillusionment with paper money. Most of the pre-Napoleonic public banks had been subject to either closure or heavy restructuring, with the storied exception of the Bank of England (Roberds and Velde 2016c, 42-43). Although great quantities of public banknotes had been issued over the preceding century, the majority of these issues were unsuccessful at either the microeconomic or macroeconomic levels (cf. Table 1). Public distrust of banknotes ran wide and deep. “Such a paper, in place of pearls and gold,” sniffs the Devil in Goethe’s play Faust (Part II, published 1832), “is convenient, as long as one knows what one has” (Bernays 1839, 40). Audiences got the joke.

Banknotes’ deservedly poor reputation did not mean that they would now be going away. Even the Devil had to admit that paper currency was convenient, and governments had not lost sight of the fiscal advantages conferred by note issue. Soon after the Treaty of Vienna, a number of newly chartered or rejuvenated public banks were again engaging in banknote issue, often quite tentatively given earlier unsuccessful experiences. Private banknotes of variable quality were also issued in some countries (e.g., England and the United States).

By 1880, this situation had totally changed. Many countries had adopted a gold standard, with silver playing a subsidiary role. Bearer banknotes, payable on demand in gold, formed a major component of the money stock in these countries. Note issue was limited to either a central bank or, in countries such as the United States, to private issues fully backed by government bonds. In place of the post-Napoleonic jumble of diffuse monetary standards and shaky paper market prices (Nogues-Marco 2013). This misalignment caused silver to be exported from Great Britain to the Continent and gold to flow in the opposite direction. In 1717, Sir Isaac Newton recommended a revaluation of the sovereign to market prices, but this recommendation failed due to political resistance, ensuring that Britain would go on a de facto gold standard (ibid., 462-463). Britain’s gold standard became official in 1816.

21 Countries where hope for public banknotes triumphed over recent experience included Austria, France, Prussia, Sweden, and the United States.

22 The United States chartered another note-issuing public bank, the Second Bank of the United States, in 1816 (Knodell 2017). The Second Bank operated a multi-state branch network and its notes circulated widely. These did not achieve full NQA status, in part due to the Bank’s policy of not redeeming notes issued by one branch at its other branches. Competitors’ resentment of this and other practices by the Second Bank, as well as President Andrew Jackson’s populist antipathy to a federally chartered bank of issue, led to non-renewal of the Second Bank’s charter in 1836 and its dissolution in 1841. In the absence of a dominant public bank, pre-Civil War U.S. banknote issue was instead dominated by almost 2300 state-chartered commercial banks (Weber 2006). The U.S. states’ policy of “free banking” was brought to an end by the National Bank Acts of 1863 through 1866, which 1) established a national charter for banks, 2) required the banknotes of national banks to be backed by Treasury bonds, and 3) taxed the state bank notes out of existence (Champ 2007).
money, a system of stable exchange rates and abundant liquidity supported rapid growth in world output and trade.

Bordo and Redish (2016) argue that the most remarkable fact about the Classical Gold Standard (CGS), as the 1880-1913 monetary regime is now known, is that it was resulted largely from self-interested actions of individual countries, rather than from sweeping multilateral agreements. Underlying the transition to the CGS were two undeniable physical trends. World silver production contracted in the early decades of the nineteenth century due to a collapse in Latin American output, and gold production expanded rapidly from mid-century, with the discovery of new lodes in California and Australia (Soetbeer 1876, 7). For the first time in 300 years, gold exceeded silver in terms of the value of world mine output. Although additional sources of silver were eventually discovered, the newfound abundance of gold made bimetallic standards challenging.23

In 1866, four countries (Belgium, France, Italy, and Switzerland) sought to jointly commit to bimetallism through an agreement on a common monetary standard, known as the Latin Monetary Union (LMU; Redish 2000, 189-197; Bordo and Redish 2016, 600-601). The legal tender value of silver coins was limited to 50 francs in each signatory country, effectively demoting such coins to quasi-tokens. Even this compromise was soon put under pressure by the discovery of new silver lodes in the United States, pushing LMU members in the direction of a gold standard.24 Following the issue of much unbacked war currency (“greenbacks”) during the Civil War (1860-1865), the U.S. restored free minting of gold in 1873 and made greenbacks redeemable in gold, thereby establishing a gold dollar. The same year saw the introduction of the German gold mark, made possible by Prussia’s victory in the Franco-Prussian War. Payments of gold indemnities by France also undermined the ability of the Banque de France to stabilize the bimetallic ratio. These changes were underpinned by a broad shift in political views regarding

23 The conventional view that bimetallism was unstable and always on a knife-edge (Jevons 1884) was challenged by Friedman (1992 chapter 6), Oppers (1996) and Flandreau (1996), who argued that before the Franco Prussian War, the Banque de France had large enough gold and silver reserves to keep the market ratio of silver to gold close to the mint ratio. France before 1870 was the largest and wealthiest country in Europe and was the center of a multi-country optimum currency area.

24 Flandreau (1996) however rejects mechanistic explanations of the LMU transition from bimetallism to gold, and instead views this shift more as a consequence of France’s defeat in the Franco Prussian War.
silver coinage and bimetallism, in places such as the new powerhouse of Germany, European countries tied to Germany (Gallarotti 1995), and the U.S. (Friedman 1992, chapter 3).

The upshot of these developments was that by 1880, gold had become the “new silver” of world trade. With the world’s largest economies on a gold standard, other countries found it advantageous to tie their currencies to gold. A country’s commitment to a gold standard ensured a stable foreign exchange value for that currency and implied a degree of fiscal commitment, albeit one that could be relaxed during times of war (Bordo and Kydland 1995). The degree of commitment varied by country, however (Bordo and Rockoff 1996). Arithmetically, the CGS was sustained by discovery of large new sources of gold, as would have been predicted by the classical commodity theory of money (Barro 1979, Bordo 1981, Rockoff 1983). Nonetheless, secular deflation occurred from the 1870s to 1896, as robust growth in the world’s real economy outpaced increases in the stock of monetary gold. The rising real price of gold provided incentives to improve refining technology and to explore for new sources, leading to the development of the cyanide refining process and to gold discoveries in South Africa and Alaska. These last developments reversed global deflation (Bordo, Landon-Lane, and Redish 2009).

In many countries, banknote issuance became a monopoly of a privileged public bank (Gorton and Zhang 2022, 3). These institutions, so often associated with monetary mischief during the eighteenth century, became symbols of national monetary virtue. They were increasingly called “central banks,” although many were still privately owned, and their banknotes were accorded legal tender status. With new privileges and new titles came new responsibilities, as central banks were called upon to provide emergency lending during financial crises (Bordo and Redish 2016, 601-602). In addition, exchange rates were not passively maintained, but were managed by central banks within narrow bands known as “gold points” (Bordo and MacDonald 1995). Central banks also exerted some influence over interest rates within their own countries, by using their balance sheets to attempt to limit the extent of shocks originating from other countries within the CGS (Bazot, Morvs, and Monnet 2022). There were limits to this mission creep, however. During this period, central banks had the technical ability to conduct macroeconomic stabilization policy but usually chose not to engage in such policies.
Exchange rate stability and the availability of emergency lending supported the growth of commercial banks. Deposits at these institutions became increasingly money-like due to more widespread use of checks (in English-speaking countries) and giro payments (in many countries in continental Europe). National banking systems were knit together through the use of specialized bills of exchange known as bankers’ acceptances, especially bills drawn on London (Accominotti, Ugolini, and Lucena-Piquero 2021), and through new communications technologies such as transoceanic cables. The end result of these developments, by the close of the CGS in 1913, was an equilibrium that resembled a fast-moving, globalized game of musical chairs, one in which the world stock of monetary claims redeemable for gold exceeded the world stock of monetary gold by a factor of ten (Triffin 1997).

Despite this apparent fragility, national commitments to gold standards were seen as credible and the overall degree of satisfaction with this system was high. This was in part because hard-won confidence in the micro side of money—as embodied in the now marvelously uniform national coinage, widely circulating public banknotes, and the free flow of capital—dominated any desire to actively manage the macro side. The predominance of micro was possible because the political economy of the CGS was based on a limited franchise. When confronted with macro shocks, countries were subject to the largely unbuffered, highly procyclical effects of the price-specie-flow adjustment mechanism. Negative shocks combined with balance-of-payment deficits led to gold outflows, deflation, and depression. Consequent rises in European unemployment were handled through a variety of nonmonetary channels, including emigration to the New World, transportation to Australia, the workhouse, and last but not always least, imprisonment.

However, the Neoclassical economists Marshall, Wicksell and Fisher, had issues with the gold standard’s record of alternating waves of deflation and inflation. They proposed plans for monetary reforms to induce greater price stability (Bordo 1993).
Transformation 4: from convertible fiduciary money to fiat money

The gold-standard music came to a stop in the summer of 1914, and would never fully resume. The carefully tended monetary standards of the CGS were abandoned during World War I, as nations confronted the fiscal realities of extended military conflict. War expenses were funded through inflation: either directly through money creation, or indirectly through the issue of nominal bonds whose value could be inflated away.

The next 30 years would provide numerous illustrations of the monetary paradox. On the one hand, by 1918 there was a great abundance of paper assets, in the form of banknotes and transactable bank deposits, which people now accepted and used as familiar forms of money. On the other hand, there was a deficit of ideas as to how the value of paper money could be credibly maintained. To the last point, there was an underappreciation of the increased potential for multiple monetary equilibria (Bernanke 1995), i.e., for the fact that the existing stock of monetary gold could support either large stocks of paper money and stable prices—as occurred under the CGS, only now there was much more paper—or experience a rapid contraction of paper money accompanied by rapidly deflating prices—as would occur during the Great Depression. Instead, the prevailing policy advice, as typified by the famous interim report of the Cunliffe Committee, was a mild update of that offered by Guillaume de Soterel in 1340: for money to function as money, prewar monetary standards would need to be restored. Anything less would constitute unacceptable cheating.

This line of thinking failed to recognize that the world had changed since 1340, in at least three major ways. The first was that twentieth-century states were no longer operating as medieval semi-autarkies but were now bound together by extensive webs of trade and finance. Returning monetary standards to anywhere close to the CGS equilibrium would thus require a high degree of international coordination. Second, warring states had issued large quantities of

26 The Cunliffe Committee was a committee of experts appointed by the British Parliament in 1918 and tasked with recommending measures for a transition to a peacetime monetary regime (Redish 1993, 786). A second version of the committee’s report acknowledged the possibility, however remote, of a currency not tied to gold (Bordo and Redish 2016, 603).
nominal debt (an impossibility for most medieval states) and deflationary monetary policies would only raise the real value of this debt. Third, there was, as compared with earlier eras, more widespread political resistance to deflation and its implied tax burdens, thanks to the postwar expansion of the franchise in many countries (Bordo and Redish 2016, 603).

By the late 1920s, the world’s major economies had groped their way to a compromise between prewar monetary virtue and postwar fiscal necessity, in a monetary regime known as the interwar gold exchange standard (Bordo and James 2014). The general idea was that each currency would be anchored to a gold value, but with a longer chain as dictated by individual circumstances (Redish 1993, 788). More precisely, in each country other than the central reserve countries of the U.S. and U.K., member countries would hold foreign exchange (pounds and dollars) as a substitute for gold reserves. This meant that their money-to-gold ratio increased greatly compared to pre-WWI. Implementations of the exchange standard varied by country. The United States, flush with the majority of the world’s monetary gold, never deviated from pre-war convertibility, a policy that exerted deflationary pressure on other countries. Most other countries adapted by limiting their citizens’ ability to transact in gold coins. International cooperation was constrained by the United States’ unwillingness to participate in multilateral arrangements. Nonetheless, the U.K. was able to return to its prewar standard in 1925, and by 1928, all major industrialized countries had restored some level of convertibility. This was however a tenuous equilibrium that would only last for three years.

The quick death of the interwar gold exchange standard revealed that it had been a much more fragile construct than its prewar counterpart. Its implementation suffered from four major difficulties. The first was the adjustment problem, under which deficit countries, including the U.K. at its prewar parity, were prone to deflation and recession (Keynes 1925), while surplus countries, e.g., France, had undervalued parities and ran persistent balance of payments surpluses, absorbing gold from the deficit countries, which in turn was sterilized. A second problem was liquidity—wartime inflation greatly reduced the real price of gold and hence suppressed gold production, contributing to a global gold shortage. A third problem was the confidence problem, embodied by a fear that a speculative attack against the reserves of the U.K. could lead to global monetary collapse. Each of these first three problems was exacerbated by a fourth problem, a general loss of faith in countries’ political will to adhere to a gold peg. The
Classical Gold Standard was based on countries’ credible commitments to maintain gold convertibility in all circumstances except in wartime. The postwar extension of the franchise meant that domestic macro issues, including especially surges in unemployment, would increasingly take precedence over maintaining gold parities (Eichengreen 1992).

The interwar standard’s house of cards collapsed in 1931, beginning with failure of an Austrian bank, Creditanstalt, in May of that year. That failure sparked a sequence of financial crises that soon forced the U.K. off convertibility, and eighteen months later, the U.S. Most other countries followed (France held out until 1936). Countries did not let their fiat currencies float freely, however, but tried to maintain a “managed” gold standard with the help of heavy restrictions on domestic use and/or capital controls. Early, broad-based attempts to coordinate these managed standards, such as the 1933 World Monetary and Financial Conference, came to naught. Eventually (in 1936) the U.S., U.K., and France worked out an informal pact, the Tripartite Agreement under which all sides intervened in each other’s currency to smooth the devaluation of the French franc (Harris 2021). In addition, France promised not to unilaterally devalue (Bordo, Humpage, and Schwartz 2015).

Transformation 5: Anchors aweigh

Widespread dissatisfaction with the interwar gold exchange standard was the impetus for the formation of the postwar Bretton Woods System (BWS), which lasted from 1946 until the United States’ closing of the gold window in 1971. Under BWS, the U.S. dollar was pegged to gold, and other currencies to the dollar. The basic idea of BWS was again to keep currencies anchored to gold (however indirectly), but with yet longer and yet more flexible chains, and with more international cooperation, than under the interwar gold exchange standard. The designers of BWS recognized the political reality, however, that countries would now seek to maintain full employment through countercyclical monetary and fiscal policies, and that domestic priorities would often dominate external balance considerations. The tension between internal and external balance was to be managed by capital controls and the adjustable pegs, which could be altered in the face of supply and other shocks.
The original design of the Bretton Woods System constrained the micro functionality of money through exchange controls and other financial regulation, in deference to domestically oriented macro policy objectives, the reverse of the situation under the Classical Gold Standard. However, efforts to evade the capital controls and other restrictions led to ingenious financial innovation, and this innovation helped weaken BWS until it could no longer be sustained. The collapse of BWS from 1968, the 1971 de-anchoring of the U.S. dollar to gold, and subsequent (1973) abandonment of dollar pegs were watershed events in monetary history. There would be no more attempts to construct a global monetary system bound to precious metal. To borrow Redish’s (1993) memorable metaphor, by 1971, most of the world’s central banks chose to weigh their metal anchor and by 1973, these were steaming off in disparate directions.

Few topics in economics have been as thoroughly researched as the collapse of Bretton Woods. Redish (1993) summarized various strands of the literature (to that date) by arguing that this monumental de-anchoring can be seen as the culminating event of previous monetary transformations, each of which loosened the ties of money to precious metal. Redish proposes, as does much of the literature (e.g., Bordo 1993) that BWS, despite its emphasis on multilateral cooperation and its sophisticated design, ultimately retained many (if not all) of the shortcomings of the interwar gold exchange standard. These defects were obscured until 1959 by restricting current account convertibility through the imposition of exchange controls in most countries. Not long after these controls were removed, astute observers (e.g., Triffin 1960) began to predict its unraveling. They did not have to wait too long for the predicted collapse to occur.

Many factors contributed to the demise of the BWS but the ones emphasized by Redish (1993, 790) bear repeating here. One factor was the changing mission of central banks, virtually all of which became nationalized after World War II and were now assigned domestic macroeconomic objectives that could conflict with a currency peg. A second and related factor was increasing political acceptance of currency devaluations, which undermined the rationale for

---

27 For example, a Google Scholar search on the terms “Bretton Woods System” and “collapse” returns 31,400 works as of this writing (September 8, 2023).

28 Triffin recognized that BWS was unsustainable. He worried that as soon as current account convertibility would be restored that the rest of the world, as they grew, would increasingly use dollars as reserves, eventually leading to the exhaustion of US gold reserves backing the outstanding dollars and precipitating a repeat of 1931 (Bordo and McCauley 2019).
the pegs. A third factor was capital mobility and financial innovation. The development of the Eurodollar market in the 1960s meant that U.S. dollars could be easily created and transacted outside of the United States.\textsuperscript{29} The roughly simultaneous emergence of various forms of shadow banking meant that dollars could be created and transacted outside the regulated banking system.\textsuperscript{30} Neither Eurodollars nor shadow dollars existed within the regulatory safety net of the U.S. banking system, but these were nonetheless seen as “sunny-day” NQA assets (Mehrling 2010, 111). Rapid postwar economic growth meant that all dollars, traditional and nontraditional, weighed on the United States’ limited stock of monetary gold. A fourth factor was the familiar bane of central banks, fiscal pressure. In the mid-1960s, the United States entered a major war and expanded its social safety net, with neither activity funded by new taxes. The Fed’s accommodation of U.S. deficit finance meant that the dollar creation machine kept running at full tilt, until Bretton Woods was irretrievably broken.\textsuperscript{31}

The decade after 1973 was one of monetary disorder. Central banks, no longer tied to the anchor of gold pegs and mandated to pursue countercyclical policy goals, attached primary weight to maintaining full employment, and attempted to do so by exploiting the Phillips Curve. Adherence to this policy approach led to the Great Inflation, defined as the period from 1965 to 1983 when the Fed and other central banks were reluctant to tighten monetary policy sufficiently to completely offset inflation. Successive failures to control inflation increased market expectations of inflation, worsening the problem. Inflation further accelerated when central banks tried to counteract the effects of oil price shocks in 1973 and again in 1979 through accommodative policies. The end result was a situation not usually observed in peacetime, of double-digit inflation accompanied by high unemployment. This was followed in the early 1980s by a wrenching disinflation and even higher unemployment as central banks tried to undo the earlier damage. Exchange rate volatility also increased as the BWS pegs were abandoned, transmitting terms-of-trade shocks to many economies.

\textsuperscript{29} A Eurodollar is a U.S. dollar deposit at a bank chartered outside the United States.

\textsuperscript{30} \textit{Shadow banks} is catchall term for institutions that functionally resemble banks but 1) have no bank charter so cannot accept deposits, and 2) have no public-sector credit guarantees (Pozsar et al. 2010).

\textsuperscript{31} Fed policy pushed real returns on U.S. debt close to zero during the Vietnam War (1964-1973), helping to finance the war; see Hall and Sargent (2021). U.S. monetary policy during this period was not as accommodative as policies pursued during the two world wars, but nonetheless fostered inflation and exerted fatal stress on the BWS.
An unplanned consequence of this failure of monetary policy was a resurgence of innovation in the micro dimension of money. As the value of money became less reliable, people devised new, less taxing ways to transact. Improvements in information technology allowed for new forms of payment and high inflation rates gave people an incentive to substitute away from cash or non-interest-bearing bank accounts. At the retail level, there was more use of credit and debit cards for point-of-sale transactions and more use of automated clearinghouse technologies for wage and bill payments. At the wholesale level, expanded large-value interbank payment systems enabled the clearing and settlement of record volumes of financial-market and foreign-exchange transactions. Much of this trading was undertaken to allocate risks associated with high inflation and volatile exchange rates.

By the 1990s, central banks had recognized the limitations of unrestrained active monetary policies and were searching for new frameworks to anchor policy decisions. In the European Union, this quest led to the formation of the European Monetary Union in 1990 and the launch of a jointly managed, multinational currency (the Euro) in 1999. More generally, central banks began to adopt inflation targets, both formal and informal, as a sort of modern implementation of De Soterel’s “middle sort of money.” The idea was that inflation should run just fast enough—a common target being two percent—to encourage firms to adjust prices to changing economic conditions, but not so fast that too many questions were asked about this price trend. This new central-banking orthodoxy also allows for temporary deviations of inflation from target to effect countercyclical policy, so long as inflationary surges are met with sufficiently aggressive policy responses, a strategy known as the Taylor Principle (Taylor 1993, 1999). At the political level, the Taylor Principle provided a compromise between forces urging higher inflation to combat unemployment and forces urging strict adherence to inflation targets.

Central bank digital currencies: the next transformation?

The new millennium began with a seemingly stable monetary order, built around the new orthodoxy of low and stable inflation, known in the U.S. as the Volcker-Greenspan doctrine (Hetzel 2023). On the micro side, retail payments were dominated by traditional payments
instruments (cash, checks, cards, etc.) that were managed by either central banks or highly regulated commercial banks, and wholesale payments were dominated by central-bank operated interbank settlement systems. At the center of this Ptolemaic universe were central banks’ policymaking committees, who increasingly focused their attention on managing inflation within small bands around a target value. Financial regulation and monetary innovation were second-order policy concerns.

The past two decades have seen serious challenges to this comfortable monetary order, challenges emanating from both macroeconomic and microeconomic developments. On the macro side, crisis events in 2008 (the global financial crisis), 2010 (the Euro crisis), and 2020 (the Covid-19 crisis) have forced central banks to push policy interest rates to the “zero lower bound” or sometimes beyond, and to provide emergency liquidity support in staggering quantities. Moreover, the new prevalence of nontraditional (shadow) banking has necessitated emergency support to a broad range of counterparties beyond the regulatory safety net. Central banks’ balance sheets, enlarged through this crisis lending, have remained large thanks to subsequent bouts of directed monetary expansion (“quantitative easing”). Financial regulations have been tightened and the financial stability mandates of central banks have been expanded. Following each crisis, there has been talk of a gradual return to the pre-2008 monetary equilibrium, but with the appearance of expanded stocks of public debt, swollen central bank balance sheets, and long stretches of historically low interest rates, such unconventional policies seem more and more like the new normal.

There is also a growing sense that the impact of these macro changes may ultimately be dwarfed by micro developments. Driving these developments is the fact that traditional forms of money and payments, while ubiquitous, trustworthy, and convenient, are both expensive and technically dominated given the capabilities of modern information technology. At the retail level, a U.S. merchant receiving a $100 dollar credit card payment from a customer may easily pay $3 in card fees. At the wholesale level, a firm supplying $100,000 of merchandise to an importer in another country may wait days for its invoice to be paid. The aggregate footprint of these frictions is large: residents of the United States and Canada, for example, are estimated to spend two percent of their GDP on payments, about $US 500 billion annually (McKinsey &
Residents of the Asia-Pacific region expend even more just to get paid, about $900 billion per year (Ibid.) Opacity is big business, and a rapidly growing one.

The gap between the immense demand for payments and the often costly and clunky capabilities of traditional payment instruments has given rise to a new class of payments instruments. The term digital currency is applied to many nontraditional forms of money and payment. The “currency” label reflects that some digital currencies, most famously Bitcoin, operate on widely accessible decentralized ledgers that can partially replicate the ubiquity and privacy of paper banknotes. While digital currencies such as Bitcoin lack the no-questions-asked quality of traditional forms of money, more serious competitors have been stablecoins such as Tether, which often promise to redeem their currencies for fixed amounts of bank money, and which purport to back their currencies one-for-one with liquid assets (Gorton and Zhang, 2023). Other, “algorithmic” stablecoins seek to maintain a constant value by offering to swap alternative digital assets for their stablecoin. The business model of stablecoins is clearly bank-like (providing putatively NQA assets that can be easily used for transactions) and their digital currencies have been compared to privately issued banknotes.

The rise of stablecoins has sparked a number of policy concerns. One is that because stablecoins are only lightly regulated, they are subject to runs, potentially leading to systemic disruptions if the stablecoin sector were to become large enough (Carapella et al. 2022). Ample historical precedents exist for this type of financial fragility (Gorton and Zhang, 2023). Recent failures of some stablecoins such as Terra’s UST have given urgency to these concerns (Wong 2022). At another extreme, widespread acceptance and advanced functionality of stablecoins could result in people substituting away from traditional bank money, leading to disintermediation and potentially, loss of monetary control for central banks (Brunnermeier, James, and Landau 2019).

Central banks have not been indifferent to these issues and have responded by either exploring or offering digital currencies (CBDCs) that can compete with stablecoins. By one recent count, the majority of central banks (131, as of July 2023) have at least considered CBDC

32 These expenditures represent transfers to the banking industry from other sectors of the economy, but are widely seen as entailing some degree of inefficiency.
issue and a few (11, as of the same date) have actually issued a CBDC.\textsuperscript{33} The CBDC label is a broad category that incorporates multiple types of instruments. A general definition of a CBDC is “a digital payments instrument, denominated in the national unit of account, that is a direct liability of the central bank” (Bank for International Settlements 2021b, 3). The Federal Reserve applies a stricter definition, under which a central-bank-issued digital currency is one that is “widely available to the general public,” i.e., a retail payments instrument (Board of Governors 2022, 5). Much of the public debate about CBDCs has emphasized their functionality at the retail level, but the impact of wholesale CBDCs may be ultimately be larger.

Proponents of CBDCs have pointed to numerous potential improvements over existing forms of money.\textsuperscript{34} One is simple technical efficiency: if CBDCs were to be implemented on decentralized ledgers, be widely accessible, and offer a high degree of privacy, these advantages could mimic the advantages of eighteenth-century paper currency over then-prevailing ledger money.\textsuperscript{35} Given the large footprint of payment systems, the impact of attendant cost savings would also be large. In addition, CBDCs would not pose the financial stability issues associated with private stablecoins. CBDCs might allow for more sophisticated, automated forms of transactions (“smart contracts”) and facilitate more imaginative forms of monetary policy, including unrestrained negative or positive interest rates on (digital) currency (Bordo and Levin 2017) and quasi-fiscal transfers (“helicopter drops”). CBDCs could facilitate cross-border transactions, particularly at the wholesale level (Bank for International Settlements 2021a). Duffie and Economy (2022) argue that the currencies of countries such as China, which have heavily invested in CBDC technology and have already launched a CBDC, will enjoy first-mover advantages in cross-border transactions as CBDC use becomes more accepted. Other central banks must accelerate their CBDC programs, Duffie and Economy argue, if they want to offer competitive cross-border functionality.


\textsuperscript{34} Prasad (2021) surveys many of the pro-CBDC arguments presented here.

\textsuperscript{35} I.e., use of a token-based rather than account-based payments instrument could reduce the information needed to validate a transaction. By analogy, most merchants are happy to accept cash but are reluctant to accept third-party checks.
The history of monetary transformations suggests that these optimistic assessments of CBDCs’ capabilities should be tempered with a dose of caution. While past centuries have seen steady progress in transactions technologies—from coins to ledger money to banknotes to checks/giro payments to payment cards, etc.—the major challenge in realizing the technologies’ benefits has been to create credible institutions to ensure that transactions occur with no questions asked. Inevitably, government has a starring role in creating such institutions, but government involvement is not a panacea. Meshing the macro and micro aspects of money is always challenging, and success may require a high degree of compromise and flexibility.

There is perhaps no historical episode that better illustrates these principles than the launch of the Bank of England (transformation 2 above). The Bank was not the first banknote issuer in Europe, so its debut did not cross any technological frontiers. The Bank was also, from its start, an institution rife with political compromise. It was chartered as a private bank, in part because no one believed that the state (neither the crown nor Parliament) could directly operate a bank without driving it into insolvency, and in part to allow the Bank to raise equity capital. The key compromise was that some portion of the Bank’s profits would have to go to its shareholders. The key advantage to the state was that these same shareholders were not expecting a speedy return of principal.

There were other limitations, as well. As would be demonstrated during the Napoleonic period, the Bank could not easily issue small-denomination notes, because the English legal system made prosecutions for fraud costly and often unsuccessful. Aware of its weaknesses, the eighteenth-century Bank chose not to shoulder the entire burden of state debt, as occurred in France under John Law. The Bank rarely held loans against real estate, in contrast to the early Riksbank, despite political pressure to expand such lending (Clapham 1945a). The Bank’s narrow and sometimes difficult path to success was guided by a sense of what was possible (Broz and Grossman 2004).

Nor was this always a clean path. The Bank’s “rough wooing” of the London financial markets left a trail of collateral damage. Knowing that its survival was not assured, the Bank

---

36 The Bank did try for a wider role in public finance, but was fortunately outbid by the ill-fated South Sea Company; see e.g., Roberds and Velde (2016b, 470).
suppressed many potential rival note-issuers, including another proposed public bank (the Land Bank), the South Sea Company, and private London banks with more than six partners. Parliament cooperated by not asking the Bank to take on debt that Parliament had no clear plan to repay. There was a general recognition that while the popularity of the Bank’s notes had given rise to a new kind of public bank, that this liquidity should not be pushed beyond its limits. As discussed earlier, this track record of successful adaptation was not replicated in many other countries in early modern Europe, where instead overconfidence in the seeming magic of banknotes led to crises that undermined confidence in the note issuers. Nor did mastery of many technical details of note issuance, as occurred with the First and Second Banks of the United States, guarantee these institutions’ political survival.

This history suggests that a range of pragmatic adaptations may be necessary in order for CBDCs to succeed. If a CBDC is intended to function primarily as a retail instrument, this will pose difficult tradeoffs regarding transactions privacy versus the potential for fraud and money laundering (Garratt 2018). If private stablecoins continue to exist alongside a retail CBDC, policy issues will arise regarding the status of the CBDC with respect to other digital currency assets (Gorton and Zhang 2022, forthcoming). The most successful type of CBDC might be one with a hybrid, public/private structure, in the spirit of the early Bank of England. At the wholesale level, a widely accessible CBDC could increase the contestability of markets for financial services, leading to fundamental changes in industry structure and posing regulatory issues (Pfister 2019). Full cross-border functionality for a CBDC might require that the central bank allow CBDC use by nondomestic parties, again raising regulatory and financial stability

37 I.e., effective anti-fraud measures can be challenging to implement in constitutional societies, as suggested by the Bank of England’s Napoleonic-era experience.

38 Nineteenth-century British experience is again suggestive of the tradeoffs involved. Banknote issue by private banks in England was originally (relatively) unconstrained outside of London. By the mid-nineteenth century, however, many in Parliament believed that the popularity of notes issued by these “country banks” posed a threat to Britain’s monetary stability. The 1844 Act that renewed the Bank of England’s charter, known as Peel’s Act, froze the note issues of the country banks and imposed a 100 percent marginal reserve requirement on notes issued by the Bank of England (Clapham 1945b, 183-5), with the intent of shoring up banknotes’ status as safe assets.

39 To this end, some CBDC proposals have envisioned “delegated” models in which digital currencies are created by central banks but are accessed through private intermediaries; see e.g., Kahn, Rivadeneyra, and Wong (2020). Delegation would thus allocate some portion of the profits from CBDC issue to private parties, an allocation that could prove politically controversial.

40 More specifically, if non-bank intermediaries could offer payment services via transfers of CBDC, this could erode the funding advantage currently enjoyed by regulated banks and contraction of the banking sector.
issues (Bank for International Settlements 2021a). Finally, a truly successful CBDC could result in a substantial expansion of a central bank’s balance sheet, with implications for the design and implementation of monetary policy (Malloy et al. 2022). It seems unlikely that any of these thorny policy issues will be resolved by mechanical refinements to the underlying technology.

We conclude our assessment of CBDCs by giving David Ricardo his due. If central bank digital currencies can offer the world a ubiquitous, technically superior way to transact in a safe and liquid asset, they represent an improvement that should and will be adopted. The history of monetary transformations shows that good ideas can overcome bad implementations. But history also validates the insight of Milton Friedman, that the greatest danger of CBDCs may lie in their success, if potential microeconomic efficiencies not balanced by macroeconomic frameworks to prevent fiscal abuse. The monetary paradox will not go away anytime soon.

---

41 For example, use of a CBDC by nondomestic parties could raise the issue of whether a central bank’s responsibility as lender of last resort would extend to such parties.

42 As with early implementations of paper banknotes, expanded balance sheets could increase central banks’ fiscal capacity and thereby the temptation for fiscal authorities to over-rely on this capacity.
References


