Central Banks and Payment Systems: The Evolving Trade-off between Cost and Risk[1]

Charles Kahn (University of Illinois at Urbana-Champaign)
Stephen Quinn (Texas Christian University)
Will Roberds (Federal Reserve Bank of Atlanta)

In Preparation for
Norges Bank conference

Of the Uses of Central Banks: Lessons from History

Oslo, 5-6 June 2014

[1]We thank participants in the Bank of Norway's Pre-Conference at the Graduate Institute, Geneva, 25-26 April 2013, for many valuable comments and suggestions. The opinions expressed here are the authors' own and do not reflect those of the Federal Reserve Bank of Atlanta or the Federal Reserve System.

Central Banks and Payment Systems:

The Evolving Trade-off between Cost and Risk

Central banks and payment systems evolved together. Many early central banks were founded as payments institutions: examples include Barcelona's 1401 *Taula di Canvi* (Usher 1934), Genoa's 1408 *Banco di San Giorgio* (Sieveking 1934a), Venice's 1587 *Banco di Rialto* (Luzzatto 1934), the Bank of Amsterdam in 1609 (Van Dillen 1934), the Bank of Hamburg in 1619 (Sieveking 1934b), and Nuremberg's 1621 *Banco Publico* (Denzel 2012). While some central banks were initially established as government fiscal agents (most famously, the Bank of England in 1694; see Clapham 1944), in most cases these institutions were soon drawn into a payments role (Roberds and Velde, 2014).

Today, payment systems continue to be a key part of central banking, and central banking remains at the center of payments. Private payment systems are important throughout Europe and North America. Innovative private systems are ubiquitous, from systems for small retail payments, such as PayPal or Square, through large value systems like CHIPS and EURO1, and up to the international CLS system. But central bank systems—Fedwire, TARGET, CHAPS, and so on—continue to be the backbone for the rest of payments.

The importance of payments activity has expanded dramatically since the 1970s with the growth of financial markets, especially the growth in foreign exchange trading post-Bretton Woods. Figure 1 summarizes the historical evolution of "payments intensity" for selected countries, measured as annual value transferred over wholesale systems relative to nominal GDP. Payments activity at the 18th-century Bank of Amsterdam was already about 2.5 times contemporaneous Dutch GDP. This ratio did not change greatly over the next three centuries: by 1960 the U.S. was turning over 4.5 times its GDP through the Federal Reserve's wholesale system (Fedwire). Post-Bretton Woods this ratio increased rapidly in most developed countries, but by 2012 appears to have leveled out at about 90-100 times

GDP, at least for the time being.[2] Payments intensity is still increasing in other parts of the world, e.g., China's ratio went from 20 times GDP in 2008 to 34 times GDP in 2012.

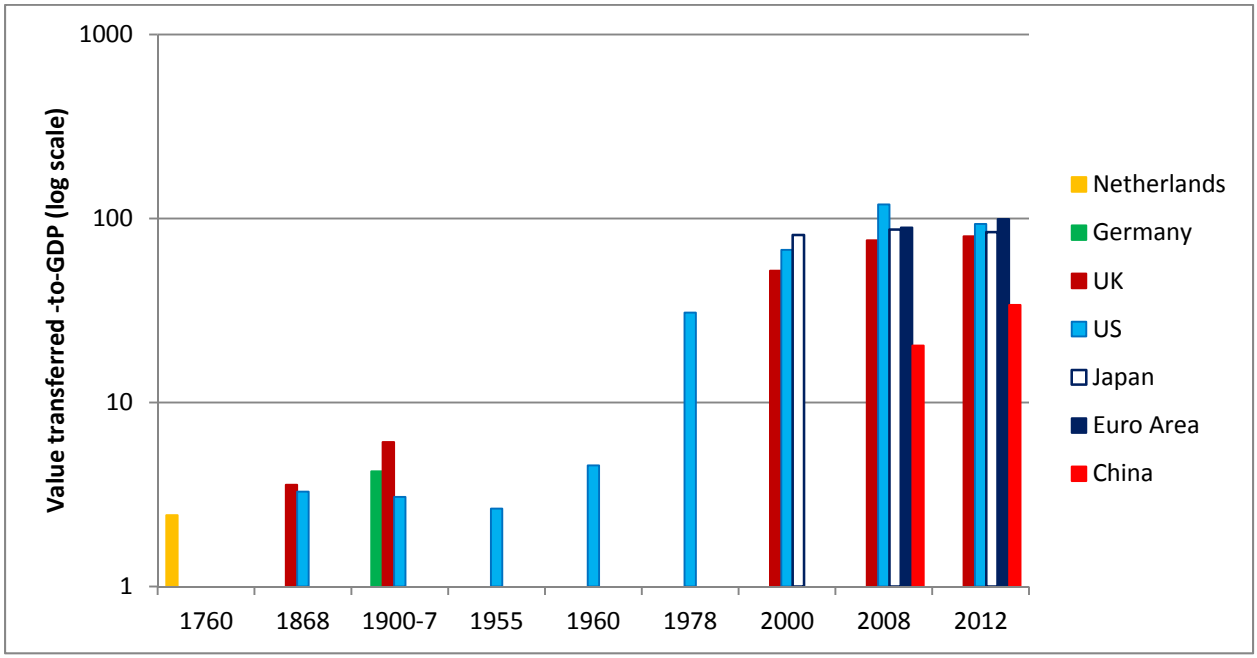

Figure 1: Large value payments to GDP ratios, selected economies

Notes: Ratios represent sum of annual value transferred over all large-value systems for a given country or currency, divided by annual GDP. Sources are Cannon 1910, Carter et al. 2006, Committee on Payment and Settlement Systems 1980, 2002, 2013, Dehing 2012, De Vries and Van der Woude 1997, Hills, Thomas, and Dimsdale 2010, Matthews 1921, Riesser 1911, Ritschl and Spoerer 1997, and Stähler 1909. Figures for 2008 and 2012 include prorated shares of CLS activity. Pre-1955 values are highly approximate.

The dramatic expansion in payments activity has created new worries for policymakers. System-wide disasters are of course of great concern. The experience of Fedwire during the events of 9/11 has led systems to pay increased attention to backup and recovery facilities. The experience of individual payment failures in large value systems and the potential for knock-on effects have led to large-scale reforms, culminating in movement to gross settlement (Bech and Hobijn 2007), the introduction of liquidity-saving mechanisms (i.e., queu-

[2] In the U.S. case, about 60 percent of wholesale payments (by value) can be directly attributed to settlement of foreign exchange trades, since they take place over systems (CHIPS, CLS) that are specialized to this function. We suspect that FX has a similar share of large-value payments in other countries.

ing schemes; see Martin and McAndrews 2008), and development of CLS (*C*ontinuous *L*inked *S*ettlement; see Kahn and Roberds 2001, as well as section 4 below). And on a day-to-day basis, the overlap in services provided by private and public systems leads to a persistent question for regulatory bodies: to what degree should the private systems, that simultaneously compete with and depend on the public backbone systems, be encouraged or restricted?

In this paper, we will present a simple theoretical framework to illustrate the evolution of central bank payment systems and, importantly, their interactions with private systems. Deficiencies in a payment system create opportunities for a central bank to improve efficiency by offering a privileged form of money. Successfully introducing central bank money then causes the payment system to adjust to its new settlement anchor. Central bank money contributes to the effectiveness of the wider payment system and its characteristics depend on the structure of the central bank. Furthermore, this co-evolution of the elements of a domestic payment system is sensitive to the pressures and opportunities created by international demand for its payment services.

To demonstrate these dynamics, we consider examples of the development of payment systems before, during, and after the introduction of central banks. First, we examine the Early Modern system of bills of exchange prevalent on the European Continent. Next, we examine the Anglo-American experience with banknotes and checks. Finally, we consider modern wholesale payments arrangements for foreign exchange, which work through multiple central banks but do not have a unifying central bank.[3]

1. Analytical framework

In order to make a transaction, a buyer and seller must establish not only the terms of the purchase—price, quantity, quality—but also the terms of the payment: when, where, and,

[3] For other approaches linking the history of payment systems to the development of central banks, see Giannini 2011 and Norman et al. 2011.

above all, how.[4] Nowadays, transactors have a variety of payment methods available to them: cash, checks, and various payment cards and internet arrangements. But as illustrated in the following sections, economic agents in earlier centuries often faced complex menus of payment methods as well.

Choosing among the alternative means of payment involves tradeoffs. As a result, an economy uses a variety of payment methods. For example, cash has high liquidity and finality, but people resist using it because cash is expensive to acquire and protect. Credit cards, in contrast, are cheaper for consumers to use, but expensive for retailers. They are also contingent and have limited secondary market liquidity. We will call the collection of these methods at any particular time, along with their supporting infrastructures, *the payment system.*

Each method of payment has a different profile of advantages and disadvantages,[5] making it most suitable for a different segment of money demand.[6] For example, if the parties to the transaction trust each other, or if the payment is relatively small, they might prefer a technique with higher risks but lower costs. As the costs of particular payments methods change, those payments methods become larger or smaller parts of the overall system.

Conceptually, sources of payments friction can be assigned into two broad camps: *resource costs* and *risk of use*. Payment instruments that have no relative advantage in either resource cost or risk are shunned, and the monies people do use have a relative advantage in

[4] Integrating the multiple dimensions of transactions into an Arrow-Debreu context presents serious challenges. One way of solving this problem is illustrated in Geanakoplos 2009, which treats each different set of terms for a purchase (in this case the collateral requirements) as a different Arrow-Debreu commodity.

[5] A host of recent research has investigated the considerations that lead individuals to choose one means of payment over another in particular transactions. See, for example, Arango and Welte 2012; Foster, Meijer, Schuh and Zabek, 2011; Kahn and Liñares-Zegarra 2012; Klee 2008; Kosse 2012; Leinonen 2008; Schuh, and Stavins 2010.

[6] The term "money" refers to a liquid asset that serves in multiple roles, the most important of them being a means of payment. Most means of payment can be classified as monies. Usefulness as a means of payment is a primary driver of demand for money (the so-called "transactions motive"), although other considerations ("speculative" and "precautionary" motives) also influence demand for money.

one or the other dimension. The set of payment technologies actually used thus exhibits the tradeoff between cost and risk (Berger, Hancock, and Marquardt 1996).

Resource costs include costs of record keeping in accounts-based payments arrangements and cost of verification in store-of-value arrangements (see Kahn and Roberds 2009). But the most important resource costs can often be summarized by the cost of the collateral tied up in the operation of the payment method.[7] There are several dimensions to the risks in using a payment system, but today, the most relevant are *liquidity risks*[8] and risks associated with *failure of settlement finality*. Historically the risk of loss of value, through inflation or outright default, was also an extremely important consideration when a transactor adopted a payments method. This risk is not a major concern for participants in established systems in developed countries today, but the recent experiences of hyperinflation in Zimbabwe, as well as persistent high rates of inflation in other developing economies, remind us that these concerns are ongoing in some payment systems.[9]

Evolution of the payment system occurs when a technological or institutional innovation reduces the costs or risks of using a payment method. Increased demand for the improved arrangement allows the innovators to earn profits. Figure 2 puts this into a schematic, where payment system evolution is that movement of a risk/cost frontier towards a zero cost, zero risk origin, rendering the old frontier feasible but inferior.

[7] In historical contexts the cost of the collateral backing the payment system (sometimes the cost of specie) is absolutely clear as will be seen below. In practical contexts it is also clear that the disadvantages of some modern systems stem from the amount of collateral or of central bank funds needed to run them (Martin and McAndrews 2008).

[8] Recent approaches to modeling liquidity risk include Holmström and Tirole 2011 and Brunnermeier and Pedersen 2009.

[9] And despite the remoteness of the risk, the possibility of default by large financial institutions and associated international payment disruption (so-called "Herrstat Risk") was the underlying driver in the development of CLS.

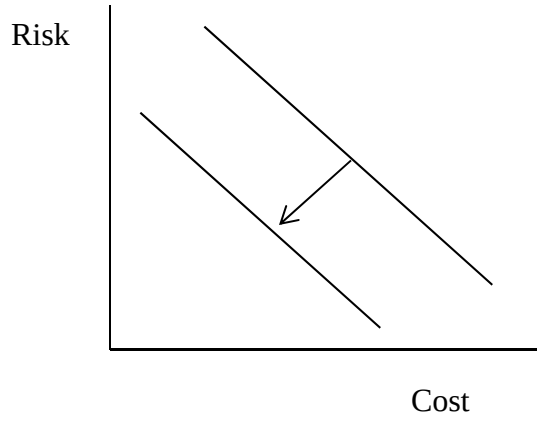

Figure 2. Payment system evolution

Source: Adaptation of Berger, Hancock and Marquardt 1996: 700.

Such evolution should not be confused with any instantaneous global jump to best practices. There are several reasons for institutions to be sluggish in reaching the technologically feasible frontier. Clearly network externalities and economies of scale are of major importance in the adoption of a particular means of payment. Thus when an incumbent is in place, entrants may not be able successfully to introduce new technologies with combinations of cost and risk that are too similar to existing systems. Instead, outsiders' innovations are more likely to arise in a different region of the efficiency frontier. Political power can also restrict the introduction of superior technologies—either through the state's use of naked power to protect its own monopoly or through influence of a powerful private system's lobbying the state. Still, over time we expect that as inferior payment instruments remain far enough behind the moving frontier, they fall into obscurity, and gradually the payment system does adjust the better to satisfy the economy's money demands.

1.1 State Money

Different types of institutions may control different parts of the payment system. At one extreme are payments arrangements run by private, for-profit corporations; at the other extreme are arrangements which are explicitly arms of the state. Most modern systems lie somewhere in between. Central banks today are state institutions, but they are typically

kept insulated from control by other parts of government. Private systems are often cooperative arrangements established by otherwise competing institutions. Typically they are charged with the dual tasks of seeking profits and providing service to their members. Even state institutions can be interested in operating payment systems so as to turn a profit.. Nonetheless, for this section we will simplify the discussion by considering the relation between a state sector providing an official means of payment and private entities competing with it.

Among competing payment systems, what distinguishes "state money," supplied by governments or their agents, from the rest? Relative to private suppliers, governments have potential "natural advantages." A sufficiently-stable government can, through its taxing authority and coercive powers, create a degree of credibility and coordination that other institutions cannot match (Kocherlakota 2001, Holmström and Tirole 2011, Chapter 5). For example, political credibility might allow a government to develop a fiat money, avoiding expensive collateral. Or a legal tender law might widely and cheaply coordinate a benchmark for debt settlement. Or government might use state power to incorporate the most reliable and stable privately-provided money available into a state money. We classify state money as *successful* when transactors choose to use it.

History shows that the success of state money is not assured. A state, or the central bank acting as its agent, might lack stability or it might lack a mechanism to confer credibility onto its money, so private arrangements may dominate. One important source of failure is a conflict between the state's short-term profits (seigniorage) and its long term goals for a payment system. The history of coinage provides many examples. For millennia, states produced coins and tried to monopolize their production. Successful mints created confidence in the intrinsic content of their coins, but many regimes gained seigniorage through the debasement of their coins. Yet other coins never became established standards, so that few of them were ever produced and little seigniorage was collected by their issuers. Other illustrations are provided by the history of central banks. Successful central banks have been able to offer a payments medium with advantages over private arrangements; nonetheless

there are many examples of institutions that either never gained traction as payment providers, or that collapsed following excessive monetary expansion.[10]

Even if not abusive, state monies may be displaced if they are inferior to the competition. State monies compete not only with private rivals, but with the monies of other states. Historically, the most successful mints created coins that circulated around the world. Similarly, the money of a dominant central bank could attract liquidity from abroad in excess of the nation's role in international trade. Important examples from earlier eras include the British pound (Flandreau and Jobst 2005) and the Dutch guilder (Flandreau et al. 2009, Dehing 2012).

Nowadays the U.S. dollar is the prime example of this "reserve currency" status; it remains to be seen whether the Euro, or possibly the renminbi, eventually supplants the dollar in this role. If it begins to happen, we can expect that the dollar won't give in without a fight. A state has many tools at its disposal in such a struggle. It may attempt to subvert competition by setting legal restrictions that favor its own money. Such legal tender rules can strengthen a currency. Promoting usage reinforces network externalities: as a particular type of money becomes more popular, the marginal benefits of holding it increase. On the other hand, efforts to impose an inferior type of money can degrade an entire payment system. Here, a relevant asymmetry is that it is usually easier to impose legal restrictions on centralized systems, so legal tender will have greater effect on debt settlement (when economically centralized through clearing operations and legally centralized through contract enforcement) than on decentralized spot transactions. Otherwise put, it is easier to use illegal money in a side-alley purchase than in a clearinghouse. Nonetheless, legal restrictions, if sufficiently severe, can even push clearing arrangements into the shadows—or nowadays,

[10] Early (pre-Napoleonic) examples of public bank failures or collapse include Genoa in 1444 (Sieveking 1934a), Venice in 1638 (Luzatto 1934), Stockholm in 1664 (Heckscher 1934), Vienna in 1705 (Bidermann 1859), and the 1720 breakdown of John Law's System in France (Velde 2007). The Napoleonic era saw the collapse of many public banks, e.g., in Amsterdam (Quinn and Roberds 2014) and again in Vienna (Raudnitz 1917). More recent examples of hyperinflation-induced collapse are (sadly) too numerous to list here: see Siklos 1995 for a survey.

out of the jurisdiction entirely and into foreign control. Access to private means of payment constrains a state's ability to impose costly public payment systems, and thus its ability to conduct restrictive monetary policies (Kahn 2013). Similarly, the state's powers in the monetary and payments arena limit the kinds of private arrangements that can develop.

1.2 Anchor

Nonetheless, the relation between the private and public spheres of payment is not simply competition between substitutes. If the public authority provides an adequate anchor, then a private system can develop from it. History provides examples of successful coins becoming payment system anchors. For a Renaissance or Early Modern city, coins (and the city's regulations regarding those coins) were the standard of finality and liquidity. Innovators responded by developing alternative payment systems that reduced costs relative to coins: mercantile credit, bills of exchange, and bank accounts. These technologies deferred the need for coin. Additional innovations avoided the use of coin through netting. Bankers learned to clear offsetting claims and merchants learned to clear offsetting bills of exchange (Velde 2009, Börner and Hatfield 2012). Eventually, multilateral netting further avoided usage of coin, so bankers centralized with clearinghouses (for their development in the U.S. and in the U.K., see Cannon 1910 and Matthews 1921 respectively) and merchants centralized with fairs. Innovation meant that the anchor, coins, moved less and less. But each innovation depended on the stable anchor.

Like coin, successful central bank money can anchor a payment system. Unlike coin, central bank money does not contain intrinsic value—it is not itself made of gold or silver. Rather, central bank money derives value from its backing—be it precious metal, sovereign debt, or the state's full faith and credit. Compared to a system anchored by coin, a central bank can reduce or eliminate usage of coin. Displacing a commodity-money anchor, however, creates new challenges for the establishment of commitment mechanisms. Again, such efforts can fail, but when a central bank succeeds, private innovators must find their spot on the efficient frontier. Relative to successful state money, private innovators can either lower costs (at the expense of risk) or lower risk (at the expense of cost). As a consequence, a new and

successful state money can set off rapid innovation—a "punctuated evolution"—as the private side of the payment system responds to the new anchor.

In fact, the private system usually directs its efforts towards cost reduction. When well deployed, the natural advantages of the public provider make it particularly challenging for private arrangements to offer a lower risk profile. This is somewhat paradoxical—after all, as we have seen, the state system has the power to renege on its promises in so many ways. But precisely because of that, a successful state system must develop strong assurance of controls on its growth—a high degree of commitment. The success of state money usually relies on credible limits on supply, and a limited supply increases the costs of this most useful resource.

As a result, confidence in an immediate means of payment has generally required assurance of some controls on its growth. But the necessary commitment makes such systems intrinsically inflexible. In the case of metals, the inflexibility was compounded by dependence on the vagaries of discovery. But more fundamentally, and particularly in fiat systems, the assurance was dependent on a belief that the rules of the game were difficult to change.

On the other hand, this inflexibility means that it is hard to improve on the backing of a stable government in periods of economic stress. The public system is likely to be most expensive but most reliable, thus serving as a refuge in times of crisis. Indeed, the contrast between the need for commitment within the central system and the need for flexibility within the economy as demand for payment grows is the tension which provides space for private systems to develop and compete. The resulting opportunity for private innovation is to offer payment services at a lower cost (but at higher risk) that many transactors find desirable. Figure 2 gives a schematic view of the process. To begin, a new state money moves inward the high cost, low risk end of the payment system frontier. Then innovation grows a new frontier towards lower costs/ higher risks. The new private system builds on the stability of the anchor.

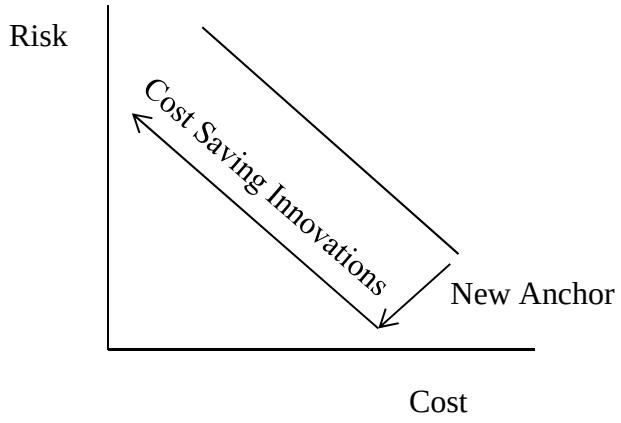

Figure 3. Punctuated evolution

1.3 System Risk

Participants choose the payments method that fits best with their preferences for mixing costs and risk. But the social costs of risk may be greater than the private costs. There can be externalities associated with the use of a payment system: misuse or failure of a payment system imposing costs on agents in the economy beyond the participants in the particular transaction. Systemic risk is inherent in any payment system: like national security, the very existence of a payment system enables the economy to rely on it to get things done, and therefore encourages production and investment; its disappearance damages everyone. More narrowly, the use of a payment system requires buying into its specific arrangements. There is value tied up in this, and so the destruction or degradation of the system causes losses to other participants in the system: the more widespread its use the greater these costs. The provider of a system will internalize these values in determining the right level of safety in order to maximize the value for the membership in its payments community, for example, through its specification of amounts of collateral to be posted by participants. To the extent that there are spillovers to non-participants, or to the extent that limited liability on the part of the system provider leaves him unaffected by systemic losses, the state may demand a higher level of collateral than even the system operator would prescribe.

The other half of this tradeoff, the cost of the protective collateral, is also the potential source of a wedge between private and social costs. The costs of providing collateral are real enough to the participants who are required to bear them. But central banks have natural advantages in the creation of reserves which can be used as backing in payment systems. A fundamental puzzle in monetary theory is the extent to which costs of central bank reserves used as collateral are truly social costs.[11]

So far we have described a situation in which there is a strict distinction between the backbone public payment system and peripheral private systems. This is an oversimplification in several important respects. First, over time, as the peripheral systems become more central to the economy, the government will extend its rule-making powers to cover them as well. Reserves to back bank notes or deposits become not just a matter of the bank's desire to maintain its business, but also a requirement of public policy—sometimes, as pointed out by Giannini (2011), under pressure from the more reliable among the peripheral providers, in their quest for quality control. Moreover, as the peripheral systems centralize, the central authority tends to provide its capital to them as well. In part this makes perfectly good sense economically: the center is the low-cost provider of reputational capital and it values the preservation of peripheral systems, at least under some circumstances. There are two limits to this process: moral hazard and sheer size. The moral hazard dimension long been recognized, but the size problem has become important in recent years, as in Iceland, for example, where the peripheral system became so large as to swamp even the sovereign's reputational capital. Finally of course, the decision to provide that reputational capital is only partly voluntary. ("Too big to fail" is not only a phenomenon of the current age; the Bank of Amsterdam felt compelled to lend to the Dutch East India Company despite prohibitions in its charter; see Uittenbogaard 2009). And so the need for rule making by the center is in part a defense against its inability to refuse to bail out private institutions.

[11] The presumed power of central authorities to provide real money balances costlessly underpins much of the debate about optimal money supplies and the Friedman rule. See Lagos and Wright 2005.

1.4 The Role of Information

Information is central to the working of payment systems. As emphasized in Kahn and Roberds (2008), the success of a system requires the ability of participants to distinguish legitimate from counterfeit tokens in "store-of-value" systems and the ability to distinguish identities of counterparties in "account-based" systems. More basically, it requires the ability to distinguish one payment system from its imitators: in other words successful payment systems must be "name brands." The ability to police one's brand is a crucial aspect of the necessary generation of confidence in the system. Historically, sovereigns executed counterfeiters for treason, and developed techniques and institutions for preserving the value of the coin.[12]

Private arrangements band together in guild-like organizations (think of Visa and Mastercard as their modern-day equivalents), not only in order to maintain oligopoly power against rivals, but also to set standards for safety of instruments and guarantee that the public not confuse inferior versions with their own. For both of these reasons payments organizations appeal to the sovereign for protection and exclusive powers, moving down the road from purely private to quasi-public organizations.

One advantage emphasized nowadays in "store-of-value" systems is their ability to provide *anonymity*: payments may be made successfully without disclosing the identity of the payer (Kahn, McAndrews, and Roberds 2005). While this side-benefit has become of increasing interest in recent years with the ever-increasing concern with privacy, this does not seem to us to have been a primary driver in the origination of any payment system before the internet era. Aside from coin, the earliest monetary instrument that permitted privacy was the bearer note. The introduction of bearer notes by the Bank of England in 1694 allowed for anonymity of transactions, but early notes were used for large value, business-to-business

[12] The most famous of these is an elaborate procedure for testing a random sample of newly minted coin for weight and fineness, known in England as "the Trial of the Pyx" (Stigler 1999). This procedure was in use as early as the thirteenth century. Virtually identical procedures were applied in other countries, see, e.g., Polak 1998 for a description of its use in the seventeenth-century Netherlands.

payments (Clapham 1944: 22-3). Their primary benefit was to facilitate finality, by allowing an alternative to chains of debt transactions. In other words the important aspect of the trail of information in earlier payment systems is not that an individual did not leave any trail, but that no one needed to worry about following the trail others had left.

The other aspect of information crucial to running a successful payment system is knowledge of counterparty quality. Consistent with the difference in risk, private systems are often confined to smaller groups of participants than the public system. The risk associated with the private system can be reduced by carefully restricting membership to individuals deemed sufficiently reliable, or by limiting transactions to those counterparties whom one can monitor readily. Indeed demand for public systems with improved guarantees only arises when the extent of the community of transactors begins to exceed the confines of such groups.

1.5 Preview

In the following sections we consider several historical examples in which a central bank or central bank innovation is introduced into an existing payment system. We examine the adjustments that occur as the rest of the payment system develops around the new anchor. We also consider the verdict on the effectiveness of the innovation, as evidenced by international participation in the system.

2. Exchange banks[13]

The first generation of central banks in Continental Europe offered accounts rather than currency. With the exceptions of Naples and Genoa, the early public banks did not circulate monetary liabilities outside their bank. Instead, Barcelona, Venice, Amsterdam, Hamburg, and others offered only giro transfer within each bank. These early central banks were limited because their goal was to bolster bills of exchange: a private part of the payment sys-

[13] This section is based on Dehing (2012) and Quinn and Roberds (2009, 2012, 2014a, 2014b).

tem that moved liquidity over long distances. Exchange banks sought to replace coins as a medium of debt settlement. They did not try to displace coins from circulation as a medium of exchange. Even so, the Continental exchange banks were mostly ineffectual, or even counterproductive. An exception was in Amsterdam, where the Bank of Amsterdam did eventually innovate to create a successful anchor money for international payments.

2.1 Coins and Bills of Exchange

In the Early Modern Era, the anchor of the European payment systems was coin. Coins of the finest reputation like the Venetian ducat, the Spanish dollar, or the Dutch rixdollar circulated widely as low risk means of payment for large-value spot transactions. By "low risk" we mean that the likelihood of such coins being of a lower fineness than expected was low for international merchants and their money changers. Gandal and Sussman (1997: 444), for example, put the accuracy of touch-stone assay at around 3 percent and the accuracy of weight at ⅓ percent, so confidence in the fineness of coins was a critical competitive advantage.

Using trade coins, however, was expensive. For example, for the mid-eighteenth century, Nogues-Marco (2013: 468) calculates a two percent cost of acquiring and moving silver from London to Amsterdam: perhaps the shortest, safest and busiest international trade route in the world at the time. Costs include brokerage, loading, freight, and assay. Insurance adds another 1 to 2 percent during peace, and even more during war (Nogues-Marco 2013: 469).

To avoid such costs, merchants used bills of exchange. A bill was an "order instrument," for example an instruction by a merchant in London to a merchant in Amsterdam to pay a sum in Dutch guilders. Instead of buying and transporting coin, a merchant could spend English pounds to buy a bill drawn on Amsterdam.[14] Usually, the exchange rate within the bill deliv-

[14] This form of payment persists in the modern world. E.g., a recent *Wall Street Journal* (McMahon 2014) describes the use of bank drafts (bills drawn on commercial banks, payable at a future date) in contemporary China.

ered more Dutch guilders per English pound than could be acquired by shipping metal. The exchange rate included a charge for time, typically around ¼ percent between London and Amsterdam. Add in brokerage and postage, and total cost might reach one percent, or one-third the cost of shipping coin.

The trade-off, however, was risk. Foremost, the person supposed to pay the bill in Amsterdam might not pay. This was called a protest, and it left the creditor seeking compensation at law in Amsterdam or even back in London. Micro-level analysis of bill protest rates is very rare, but Santarosa (2010: 13) does find 44 percent of bills were protested in Marseille around 1780. London and Amsterdam protest rates were likely less, but we have no good estimate, and, as with other debts used as money, the likelihood of default would suddenly increase during a crisis. For our purposes, the relevant point is that compared to coins, bills of exchange were a high-risk, low-cost means of payment supplied by private parties. Government, however, did play a crucial role supporting this part of the payment system by enforcing bill contracts. And here is where the early public banks emerge.

Beyond assuring that contracts would be enforced expeditiously, localities sought to clarify the terms of debt settlement. Most commonly, governments would assign an ordinance value to coins denominated in the local unit of account. For example, a legal tender law might say that a particular coin is worth one guilder for settlement of debts public and private. In this way, creditors would know what coin they were due, and thus bills of exchange encouraged. Such legal restrictions could also be self-serving, for they could create demand for coins produced by local mints, and local mints paid profits from seigniorage to domestic government. To gain this advantage, however, domestic coin had to deliver more unit of account per ounce of silver (or gold) than rival coins. The ratio of value per ounce of metal is called the mint equivalent. If a coin's mint equivalent was high enough, merchants would convert bullion or foreign coin into domestic coin at the local mint (Sargent and Velde 2003).

In spot transactions, merchants could circumvent this process by valuing foreign coins more than ordinances assigned the coins (Rolnick, Velde, and Weber 1996). In debt contracts, however, debtors could insist on repayment at ordinance values. In this, debtors and

the local mint had a shared desire to create local coins that disadvantaged creditors. This dynamic was acute in the Early Modern Netherlands because a number of mints could produce legally-favored coins. The competition between mints damaged the reputation of coins by encouraging incremental debasement. Slightly less silver per coin meant a large mint equivalent ratio. In other words, legal restrictions often made local coins the anchor of the international payment system, but those same ordinances could promote the degradation of those same coins. The incentive came from an ability to shift the cost of coin debasement onto creditors, so an imperfect anchor undermined the private sector payment technology built on it. In effect, mints and debtors appropriated some of the cost savings created by bills of exchange.

2.2 Enter the public bank: the case of Amsterdam

Around 1600, Amsterdam was becoming the commercial and financial hub of northern Europe (Gelderblom 2013). The quality of Dutch coinage, however, was suffering mild debasement, and merchants in Amsterdam thought it bad for the bill business. So, in 1609, the city created a bank, the Bank of Amsterdam, whose design was based on an earlier institution in Venice. The city required that bills of exchange settle on the bank's books rather than in coin, and it pledged that at withdrawal its bank would deliver coins of a consistently high quality. The Bank of Amsterdam would protect creditors.

To do this, the bank would suffer an asymmetry: it would accept at deposit Dutch coins with slightly less silver per coin than it would subsequently give out. To prevent arbitrage, the Bank of Amsterdam charged a two percent withdrawal fee plus additional fees for coins in high demand. These fees were greater than the difference between circulating coins and the coins the Bank of Amsterdam was obliged to deliver at withdrawal. The high withdrawal fee also meant that a secondary market developed. Instead of withdrawing coin, a broker would match an existing bank customer wanting out with a prospective customer wanting in. One person would transfer money within the bank at no fee, and the other would deliver coin outside the bank at a brokerage fee less than the bank's withdrawal fee. In time, brokers became market makers ready to buy or sell at all times.

All this is an example of a new type of secondary market and private intermediary developing to lower the cost of using anchor (Bank of Amsterdam) money. Risk, of course, also went up because dealers did not assure the quality of their coinage with the same credibility as the Bank of Amsterdam. Dealers further reduced costs by offering accounts for non-bill payments outside the Bank of Amsterdam. The secondary market now swapped Bank of Amsterdam balances for private bank balances. And again, risk increased, for now customers had to worry about the private bank's liquidity in addition to the quality of coin they might eventually get at withdrawal.

The Bank of Amsterdam was itself not without risk. In concept, the Bank of Amsterdam was to be a fee-driven, full-reserves "narrow" bank. In practice, the city used its bank to lend to the city's lending bank, the Dutch East India Company, to the Province of Holland, and to important quasi-public persons such as mint masters and tax receivers. After a few decades of heavy lending, the Bank of Amsterdam learned to restrain its credit creation. This conservative position allowed the bank to survive a large run in 1672 when French troops almost overran Holland. Similarly structured public banks in other Dutch cities (Mees 1838) and Hamburg (Sieveking 1934b) did not fare nearly as well, and were forced into lengthy suspensions.

But even the Bank of Amsterdam found it difficult to flourish during the Dutch Golden Age. High withdrawal fees meant coin only infrequently left the bank, but coin deposits were even less frequent. As a result, the Bank of Amsterdam was slowly losing coin in the 1660s and 1670s. It offset the leakage with open market purchases, so the total amount of bank money remained steady. Still the demand for bank money was limited and merchants were unwilling to deposit coin at the Bank of Amsterdam for short term purposes. Coins flowed through the city of Amsterdam to the Baltic, the Mediterranean, and especially Asia, but those coins did not pass through the Bank of Amsterdam.

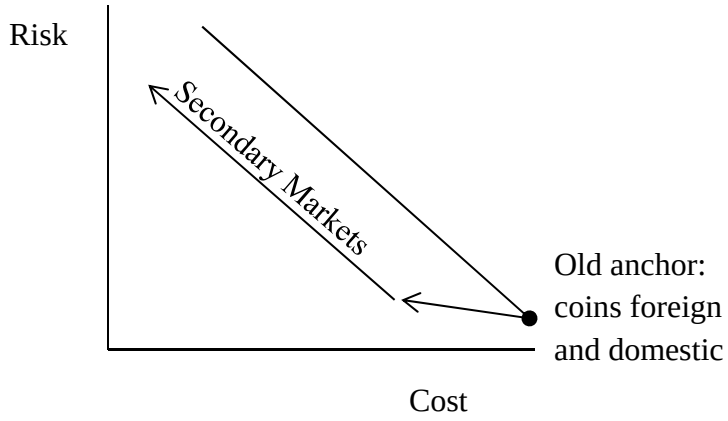

Figure 4. Adding the Bank of Amsterdam

2.3 From public bank to central bank

In response to this stagnation, the Bank of Amsterdam made a small but important change. Starting in 1683, deposits were given account balances and a receipt for the specific coins deposited. Receipts allowed the Bank of Amsterdam to separate the right of coin withdrawal from account balances. After creating that separation, the bank stripped the inherent right of withdrawal from accounts. By themselves, Bank of Amsterdam balances became a type of fiat money. This new system proved popular with Europe's merchants, and demand for bank money grew even among those not compelled by legal restrictions. Bank of Amsterdam money became the leading international currency of the Eighteenth Century, and new banking structures emerged in Amsterdam because of it.

How did the nexus of receipts and fiat money revolutionize the Bank of Amsterdam as a central bank? It lowered both costs and risk. The development of receipts made it possible to offer withdrawals at very low fees (typically ¼ percent) because customers could no longer arbitrage between types of coin. With a receipt, one got the same coins originally deposited. The bank was scrupulous in not lending these coins, and receipt commitments seem to have also deterred the city of Amsterdam from taking these coins as seigniorage.

Receipts created a credible narrow bank within the larger bank, so accounts with a receipt got lower costs and less risk.

Accounts without a receipt also benefitted. Receipts were transferable, so account holders could purchase this low-cost option from other customers instead of paying the bank the higher traditional fees. As this secondary market now served the demand for coin withdrawals, traditional withdrawal fell into disuse except, potentially, during a run on the Bank of Amsterdam. Mindful of this remaining risk, the Bank of Amsterdam quietly ended the right to withdraw accounts without a receipt. Without a receipt, bank balances became a type of fiat money. Customers could transfer them within the bank but could not compel the bank to surrender assets in exchange for them. Limiting the scale of a deposit run to the amount of coin under receipt meant that the Bank of Amsterdam could not be driven to failure. Collective action against the bank could only weaken the exchange rate; it could not force the bank to suspend payments.

The 1683 introduction of quasi-fiat money had a strong impact on the bank's payment business. Dehing (2012: 140) estimates that total "giro" turnover through the bank's accounts increased from 204 million florins in 1676 to 249 million florins in 1695. Payments through the bank increased further in the eighteenth century, reaching a peak of perhaps 400 million florins during the Seven Years' War (1756-1763).[15] As noted in the introduction, this is about 2.5 times contemporaneous Dutch GDP, a remarkable level of payments intensity for the time, equal to that attained by the U.S. roughly two centuries later (Figure 1).

The popularity of the Bank of Amsterdam's post-1683 payment regime is also reflected in the price of bank money. Figure 5 gives the fee markets charged month by month from January 1700 to January 1790. During this time, except for periods of war, the price to sell bank money (relative to circulating coin) rarely climbed over 1 percent and rarely fell below zero.

[15] Authors' extrapolation based on payments volume estimates given in Dehing (2012: 82).

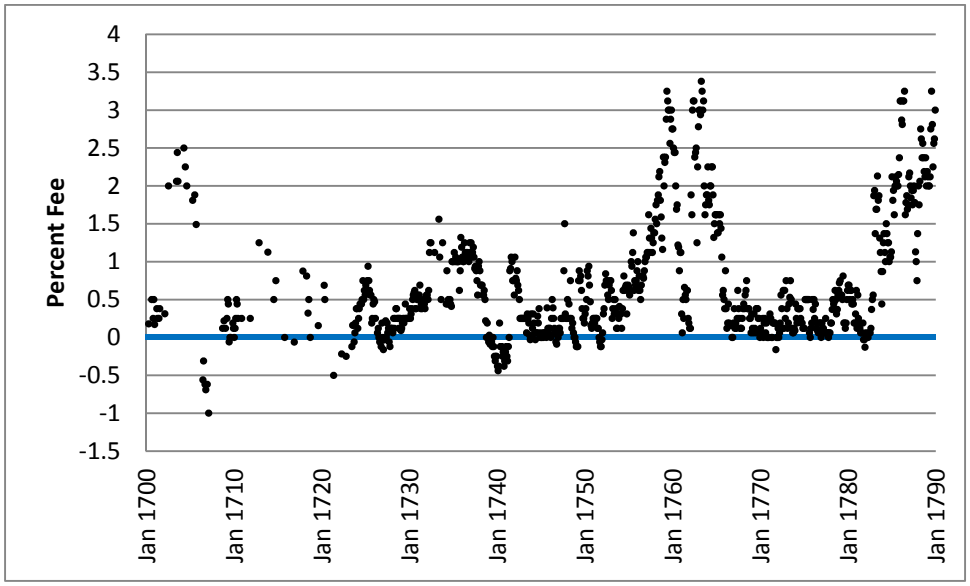

Figure 5. Domestic Market Price of Bank of Amsterdam Money

Notes: Sources are McCusker; Gillard; Amsterdam Municipal Archives. Derived by subtracting the market domestic exchange rate (current guilder/bank florin) from the deposit rate of the silver rixdollar coin (1.05 current guilder/bank florin)

This price stability, combined with Amsterdam's lack of capital controls and advanced financial markets, made Bank of Amsterdam money a successful anchor for the international payment system. Intermediaries responded by developing new types of credit systems that settled using bank money. The most important new players were merchant banks. Unlike commercial banks funded by deposits, merchant banks were funded using bills of exchange. They offered borrowers credit by accepting the bills of exchange drawn abroad (known as acceptance credit). The merchant bankers then issued new bills to fund the acceptance credit. The greatest of these firms (Hope, Pels, and Clifford) became famous in their age. These merchant banks used the Bank of Amsterdam to settle a credit network that extended to most commercial hubs in Northern Europe.

The Bank of Amsterdam's role in bank settlement also opened the opportunity to act as lender of last resort. When a major merchant bank failed in 1763, the acceptance credit market convulsed. Suddenly, banks could not sell new bills to finance bills due, so banks rushed coin to the Bank of Amsterdam to get the liquidity they needed. The Bank of Am-

sterdam even created a new liquidity facility that helped a couple of especially troubled banks. In all, the Bank of Amsterdam succeeded in saving Amsterdam's merchant banks, but it could not assist the international customers of those banks. In other words, the financial system that settled in Amsterdam extended well beyond Holland, and this mismatch limited the Bank of Amsterdam's ability to as lender of last resort.

As successful as the Bank of Amsterdam was for most of the eighteenth century, it suffered from a brittle design. Receipts created credibility but very limited fee revenue (approximately 50 basis points per year). Supplementary bank lending to the Dutch East India Company brought extra revenue but also fractional reserve risk. The Bank of Amsterdam kept such lending modest until around 1780. Dutch shipping under the flag of neutrality during the American Revolution angered Britain to the point of declaring war in 1780. The war forced the Dutch East India Company to spend heavily to arm its ships while disrupting the return of cargo from Asia. To finance this situation, the company borrowed heavily from the Bank of Amsterdam and others, but soon the company was unable to repay. The Bank of Amsterdam became insolvent. Fearing some type of default, receipt customers removed coin. The remaining customers, lacking receipts, could not withdraw coins, so the price of bank money broke trend (see Figure 5). The end of the war with Britain in 1784 did not restore the Bank of Amsterdam's credibility. Bank money endured, but it was no longer the "reserve currency" of Europe. One consequence was that merchant banks moved operations across the channel to London (Carlos and Neal 2011).

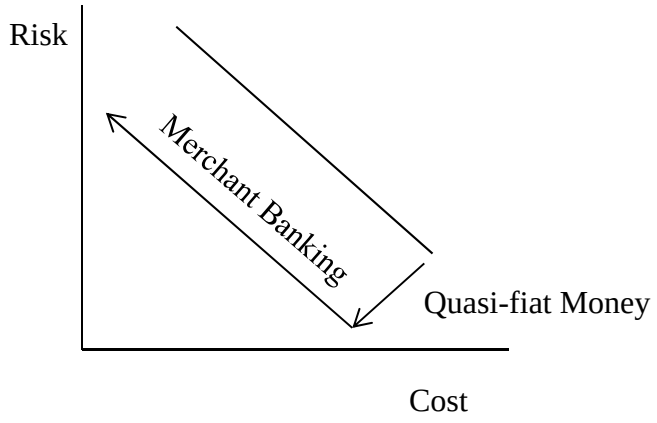

Figure 6. Mature Bank of Amsterdam

3. Anglo-American contrast

The Anglo-American evolution of central banks and payment systems took a different direction than on the Continent. Instead of municipal exchange banks, London, and then Philadelphia, focused on banks that issued currency backed by sovereign debt. Privileged note issue brought the central banks fiscal strength, yet central bank existence and independence remained a challenge to secure. And when that failed, commercial banks created quasi-central banking arrangements to support the payment system. The role of central banks in payments makes the U.S. a compelling contrast to England (James 2012b: 289-291). The two countries' payment system histories are similar enough that the differences outline the role of a central bank's money in the evolution of a payment system.

3.1 Central bank innovation: the Bank of England

Silver coins in Seventeenth Century London suffered from clipping. This created uncertainty regarding their weight or additional assay costs. Some Londoners avoided coin by adopting what was called the "banking habit." In the 1650s, goldsmith-bankers began to offer

checkable deposits for local payments and to arrange bills of exchange for international payments. Some banknotes were issued at this time, but these appear to have been a minor payment instrument (Quinn and Roberds 2003). Alternative payment services reduced costs relative to coin, but, of course, banks were subject to the risk of failure—despite being conservative fractional reserve operators by modern standards. For example, in 1685, loans comprised 42 percent of the assets of Child's bank (Quinn 1994:48), and in 1702 loans were 38 percent of the assets of Hoare's bank (Temin and Voth 2013: 67).

Early bankers also created infrastructure that further reduced the cost of payments. London bankers had bilateral clearing arrangements (Quinn 1997). At least one banker kept agents in foreign ports to facilitate reliable acceptance of bills of exchange (Neal and Quinn 2001). And the largest bankers acted as both tax collectors and sovereign creditors, so taxes due the Treasury could net debt repayments due the bankers.

Exploiting scale economies, the Bank of England's incorporation scheme of 1694 built on this infrastructure. Unlike banker-led syndicates, the corporation was able to raise large amounts of outside capital because its limited-liability stock was easily transferable. And rather than deal in large amounts of coin, the Bank of England issued large amounts of currency when lending and then accepted it back for subscription payments. The business model was a successful application of network externalities: the Bank of England made large-scale issuances of currency to acquire sovereign debt that then backed the currency. As large amounts of the currency circulated in London, expectation of acceptance became routine.

While the Bank of England's money competed with that of other banks, its favored position meant lower risk. Just two years after its founding, the recoinage of England's silver coins created a liquidity crisis and a run. The Bank of England suspended payments, and it would do so again when it was unable to meet its convertibility obligations. While not explicit in law, the Treasury granted this privilege in 1696, 1797 and 1914. While infrequently resorted to, this opt-out was important. Whereas the Bank of Amsterdam could not fail because a portion of its money was always inconvertible, the Bank of England did not fail because all of its money could become temporarily inconvertible.

The Bank of England also secured the stream of seigniorage from note issue. In 1697, the Bank of England gained a monopoly on corporate banking in England and Wales, and forgery of its notes was made a capital offense analogous to counterfeiting coins. As a result, its seigniorage from currency would suffer no large-scale threat until joint-stock banking finally emerged in 1833. Even then, the new corporate banks were kept from issuing currency if they operated in London. This fiscal strength lowered the risk of Bank of England notes, for they were backed by both sovereign debt and by the discounted present-value of currency seigniorage.

The primary risk for the early Bank of England was political. The Bank of England's charter was not perpetual, and the government repeatedly negotiated extensions when the Treasury needed new funds from the Bank of England (Broz and Grossman 2004). In effect, the state clawed back some seigniorage through new, below-market borrowing. What is remarkable, however, is how much the government did not take. The Bank of England regularly paid seigniorage profits to shareholders through dividends (Clapham 1944: 292). In contrast, the Bank of Amsterdam passed all its profits to the city, just as central banks today pay their profits to their controlling political authorities.

How the Bank of England gained secure seigniorage appears to have been something of an accident. The Bank of England's start as a corporation was a gamble at a time of intense fiscal stress on the English state. Then, the corporate form proved useful in 1697 to the state as an instrument for debt-for-equity swaps. The swaps let the Treasury convert short-term debt during a rollover crisis. Political winds, however, then blew against the Bank of England when the Tory party came to power in 1710 (Stasavage 2008: 99-129). Tory governments issued Exchequer bills that competed against banknotes, and supported the South Sea Company's gambit to displace the Bank of England in 1720 (Kleer 2012). But the collapse of the South Sea Bubble swung political support back to the Bank of England, and the mood of the era, embodied in the Bubble Act of 1720, emphasized the importance of stability (Harris 1994). The Bank of England endured as a for-profit quasi-arm of the British Treasury.

With stable political backing after 1720, Bank of England notes became the anchor of London's payment system. Again, sovereign debt and seigniorage made them low risk, but scarcity made them costly to use. Before 1760, the Bank of England focused on sovereign lending, so the supply was inelastic to aggregate demand (but elastic to war finance). Private lending was small and limited to customers who lived in London and were engaged in commerce. When the Bank of England relaxed standards enough to lend to banks (called rediscounting), the Bank of England still limited itself to buying high quality paper from the few banks that kept an account. As late as 1793, only a third of London commercial banks had balances with the Bank of England (James 2012b: 297).

3.2 Failed attempts at innovation: Banks of the United States

Against political opposition, Alexander Hamilton succeeded in chartering English-style central banking in the new United States in 1791. The First Bank of the United States was a nationally chartered, for-profit corporation whose primary asset was sovereign debt. Its primary liability was privileged banknotes. The U.S. bank had the only interstate charter while most state-chartered banks could not even open intra-state branches. Furthermore, the U.S. bank's notes were legal payments for all debts to the U.S. government.

Unlike in England, the First Bank of the United States did not have a war-time crisis with which to negotiate its first charter renewal. Instead, the war came a year after President Jefferson blocked renewal of the First Bank of the United States. Financing and supplying the War of 1812 over the length of the Atlantic seaboard convinced many, including military leaders, of the need for a central bank as an agent of the Treasury. In the meantime, the U.S. Treasury issued emergency notes, inflation surged, and state banks suspended specie redemption (Rockoff 2000: 654-5).

After the war and the election of a new president, the Second Bank of the United States was chartered for 20 years starting in 1817. The Second Bank was larger than its predecessor, but similar otherwise, and again political opposition was unrelenting. Andrew Jackson campaigned for President twice with the goal of ending the Second Bank, and, in 1832, he famously vetoed the re-charter authorization. The U.S. then entered a long period without a

central bank, and, in the 1840s, the U.S. Treasury withdrew government funds from banks and the financial system altogether. This outcome is in sharp contrast to the post-1720 political equilibrium that supported the Bank of England.

3.3 Private sector innovation in check payments

In London, as a substitute for Bank of England notes, banks offered access to bank payment services on less restrictive terms. Most lending was at, or near, usury limits, so credit rationing was the binding constraint of the era (Temin and Voth 2013: 73-94). The payment instrument of choice, however, was the check. London banks with six or fewer partners could issue notes, and some did in limited amounts, but none did in any substantial quantity. Perhaps London banks lacked the credibility to directly compete with the Bank of England, perhaps the Bank of England somehow threatened issuing banks, or perhaps most wholesale customers preferred checks. In contrast, Bank of England notes did not usually circulate outside of London, and country banks (located outside of London) issued notes for regional payments. The primary country bank payment service, however, was to supply bills payable on a London correspondent bank (James 2012a).

Because London banks used checks to lower costs, the payment system developed a thick interconnectedness. Checks gain network externalities as local banks accept checks drawn on their rivals. In the process, banks gain routine obligations on each other in the form of checks due for payment. This new system made extensive use of Bank of England notes as a settlement asset. Banks were likely settling checks bilaterally in Bank of England notes even before they created the London Clearinghouse in 1776. The clearing house adopted multilateral netting in 1841, and so reduced the amount of Bank of England notes that participants needed (James 2012a: 135). In this way, Bank of England notes became the anchor of the London banking system and, in time, the center of the English banking system. Country banks and foreign banks used London correspondent banks to secure acceptance of their bills in London, to secure access to the stock and debt markets, and to secure access to the international payments market. London clearing arrangements lowered costs and centralized risk for the nation and much of the rest of the world. When corporate banks emerged

in mid-19th century England, branch networks centered on London, and the system's reliance on the Bank of England continued.

In the U.S., the note-check divide was over time instead of over space. Before the Civil War, state-chartered banks issued banknotes, so commerce could avoid the use of coin. While U.S. banknotes were cheaper to acquire than coin, they certainly were riskier. The era famously had such a diversity of note issuing banks that entrepreneurs published guides to help merchants judge authenticity and quality, and dealers used superior information in a manner similar to coin-based moneychangers. Still, such cost- and risk-reducing operations developed because state bank notes did circulate widely. Within a city, most notes passed at par, and railroads and telegraphs reduced the discounts of notes that traveled beyond their city of origin (Gorton 1996; Jaremski 2011). Country banks set up correspondent relationships with trade-center banks (Weber 2003). In some respects, at least, the central bank anchor was not missed in its absence: Inter-city exchange fees were less after the Second Bank of the United States than under it (Bodenhorn 1992; Knodell 2003). And, as with England, the U.S. inter-regional system of notes and bills grew increasingly centered on the metropolis. By the Civil War, New York banks were the hub of inter-regional payments (James and Weiman 2011).

The National Banking Acts of 1864 and 1865 drove state banks into checking and limited the stock of notes that national banks could issue. Check use had been growing before the civil war in local, wholesale payments (James and Weiman 2010: 238). Indeed, by 1860, the level of deposits in the U.S. roughly equaled the level of bank notes, and banks in many major cities had already created clearinghouses to settle them. After the Civil War, the volume of checks continued to grow faster than notes and surged well past notes after 1890. New York was the center of settlement as banks across the US used correspondents in New York for inter-regional transfers, access to the markets, and for foreign exchange. Like London, the epicenter was the clearinghouse.

The lack of a central bank did not stop the growth of the American banking system, and the prevalence of unit-bank regulations caused that that growth to be in the number of banks.

Figure 7 gives the number of state banks, and national banks after 1863. The surge in state banks after 1880 relied on the inter-regional system of check clearing.

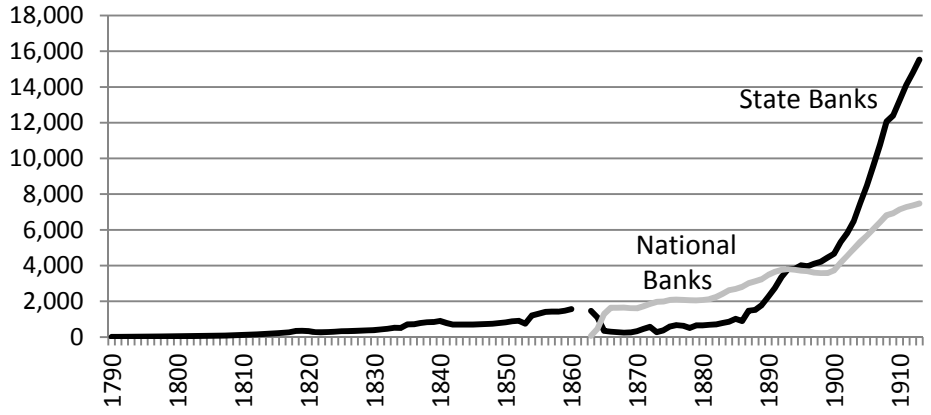

Figure 7. Number of US Banks, 1790 to 1913

Sources: Wright 2001 (1790-1820), Bodenhorn 2001 (1820-1860), Grossman 2003 (1863-1913).

3.4 Systemic Implications

Despite their differences, the nineteenth century British and American check payment systems appear to have supported comparable levels of payments activity. In 1868 (the first year for which data becomes available, since settlement occurs through Bank of England accounts rather than with notes), the London Bankers' Clearing House settled £3.4 billion in London-area payments through the Bank of England (Matthews 1921: appendix II), which is about 3.6 times contemporaneous GDP (see Figure 1). That same year, the New York Clearing House handled payments of $28.5 billion or 3.3 times GDP (Cannon 1910, 217). The British system expanded to all of England by 1907 and cleared over six times GDP, while the U.S. ratio (based on New York only) declines slightly to 3.1. However, by the early twentieth century there were over 200 regional check clearing houses operating in the U.S. (for which statistics are unavailable), so the aggregate ratio for the U.S. may be substantially higher.

Checks, bills, and their settlement infrastructure lowered costs and increased volume, but also created systemic risk. Troubles with an individual bank could spread via clearing and settlement to other banks. In this way, the supplier of money used for clearinghouse settlement gained the opportunity to act as a systemic lender of last resort. In London, this role was played by the Bank of England. In New York, the clearinghouse itself became a LOLR. And here crucial dissimilarities develop.

The Bank of England's (implicit) ability to suspend payments backstopped the system and could prevent commercial banks from suspending (James 2012b). Moreover, the Bank of England could expand lending to banks. This it did aggressively when it suited the Bank of England's operational goals, such as when convertibility was suspended (1797-1825) because of the Napoleonic Wars. Such lending, however, was limited when it went against the Bank of England's internal interests, such as during the Panic of 1825 (Neal 1998). Even when the Bank of England did clearly lend to support the system, it denied any obligation to do so (Bignon, Flandreau, and Ugolini 2012).

In New York, the clearinghouse could, and did, create emergency liquidity during crises, but the amount it could produce was limited to the collective assets of its member banks. The New York clearinghouse had no external reserves the way the Bank of England had its own holdings of gold and sovereign debt, separate from members of the London clearinghouse. As a result, when a crisis pushed the English system to its breaking point, the Bank of England could suspend convertibility into gold, so London banks did not have to suspend their convertibility into Bank of England notes. In contrast, when a crisis pushed the New York clearinghouse to suspend convertibility, it took member banks with it. "Such temporary suspensions were staple strategies of American bankers in times of crisis. … In London there was never a general suspension of payments during times of panic (James 2012b: 290)."

James, McAndrews, and Weiman (2013) argue that the U.S. system had grave macroeconomic consequences. With general suspensions, local means of payment suddenly came into short supply, so both payroll and debt servicing were imperiled for otherwise healthy firms. Also, interregional payments propagated the suspension to other cities, as respond-

ent banks had to scramble for alternative sources of liquidity or default on their own payment commitments.

How serious was this difference in deep-crisis performance? The ultimate judgment seems to rest in the foreign exchange markets. The world's money favored London, and Amsterdam before it, but not New York (Flandreau et al). Indeed, one reason the New York banks campaigned for adoption of the Federal Reserve System was to improve the dollar's international attractiveness (Broz 1999). Despite similar payment systems, how the English anchor disconnected from coin (the Bank of England suspending payments) seems an important advance relative to how the U.S.'s disconnected (general bank suspensions).

In our conceptual framework, both the UK and the US developed anchors distinct from coin, but the British anchor was less prone to suspension, and so was the resulting payment system built upon it. As a result, **figure 8** shows London's check-based payment operating with less systemic risk (at any given cost profile) than New York's. As a result, the British pound became a reserve currency, and much of the world's finance occurred in London (Flandreau and Jobst 2005: 990).

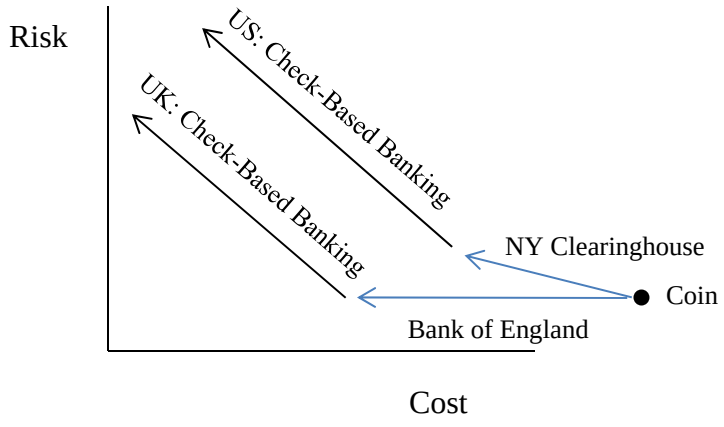

Figure 8. Anglo-American Contrast, circa 1900

When the Federal Reserve was finally created a century ago, its initial structure was designed to address both domestic and foreign payment system challenges. Foremost, the

new system held reserves distinct from those of member banks. It created a system of inter-regional check clearing that helped reduce propagation of liquidity shocks (Gilbert 2000).[16] It actively promoted the international banker's acceptances (Ferderer 2003). As a result, the dollar slowly became a world reserve currency (Eichengreen and Flandreau 2012).

4. CLS[17]

The rise of national central banking in the nineteenth and twentieth centuries did not do away with the old problem of how to move international liquidity. Under the gold standard, in principle individuals could acquire foreign exchange by redeeming local currency for gold, and then shipping the gold abroad in order to acquire foreign currency. Few did so. Instead, bankers avoided those costs through a variety of financial instruments for inter-bank transfer, such as bills of exchange and banker's acceptances.

Post-Bretton Woods, people could no longer redeem gold, and interbank financial instruments became the only method available for transmission. Meanwhile, technological advances rapidly decreased the cost of interbank transfer. Indeed, the most striking empirical regularity in payments is the worldwide increase in payments intensity since 1970 (Figure 1). Judged by this metric, the nature of the payments business has changed more during the past 44 years than during the preceding two centuries. The increase in payments intensity mirrors the increase in financial markets trading, particularly trading in the markets for foreign exchange (FX). But while the volume has changed dramatically over this period, the nature of the transactions has not. FX transactions are commonly thought of as instantaneous trades of fiat money—one central bank's liabilities against another. But, at least up until 2002, they were simply faster versions of the old interbank transfer mechanisms.

[16] Though still imperfectly, see Richardson 2007, Mitchener and Richardson 2013.

[17] The discussion of CLS here is based on Kahn and Roberds (2001). See Committee on Payment and Settlement Systems (2008) and the CLS website (www.cls-group.com) for additional information.

Since 2002, however, central banks have increasingly detached themselves from FX trades, by delegating their settlement to a private institution, the CLS Bank. Traditionally, banks used bilateral financial instruments to bridge different units of account. Now CLS can make those connections, and its account transfers replace the instruments. By operating simultaneously in multiple currencies, CLS is able to control risks of settlement in a way that no single central bank could. CLS may be the most unusual financial institution ever established. By day, it is the largest institution on the planet. By night, it hardly exists. It handles about half of the world's foreign exchange transactions, but it is also privately owned and operates with fewer than 500 employees. It was originally designed to do one job—settle international payments—and it does it with extraordinary efficiency. Pressure from the world's central banks more-or-less forced CLS into existence, but its position leads to extremely thorny policy questions for those same central banks.

In terms of our conceptual structure, CLS is a twist. From the point of view of a financial institution "paying" for the purchase of foreign exchange, CLS becomes the anchor technology: conducting the transaction through CLS is less risky (and slightly more expensive) than paying directly with central bank money. In order to economize on the collateral costs of conducting its business, CLS has incorporated a large number of collateral-saving devices (some of them inducing slight increases in the risk of the system). Finally banks have available to them bilateral transactions ("in-out swaps") officially outside of CLS, which can further reduce the collateral costs of using CLS, again with increases in the risk of delivery failure. Figure 9 illustrates this space of alternatives.

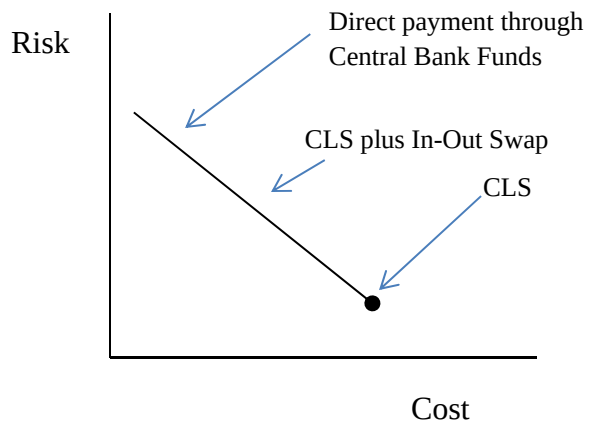

Figure 9. CLS Spanning Domestic Payment Systems

4.1 CLS, a private sector innovation

CLS came into being in 2002, as a result of regulatory dissatisfaction with traditional arrangements for settling foreign exchange transactions (Committee on Payment and Settlement Systems 1996, 1998). Settling foreign exchange (FX) trades poses special challenges both because of the sheer size of the post-Bretton Woods FX markets, and because the underlying nature of foreign exchange creates risks that are resistant to traditional risk-limiting strategies such as netting and counterparty substitution. The initial impetus behind CLS was to move the payments used to settle FX trades away from traditional large-value systems (mostly run by central banks) to a specialized institution that could better handle these risks. Although in many cases there is no legal compulsion to use CLS, it has nonetheless enjoyed considerable success. The most recent statistics available on the CLS website (as of this writing, February 2014) indicate that CLS is currently settling a little over $5 trillion daily (counting transactions on both sides) or roughly 50 percent of the world's daily FX turnover (Bech and Sobrun 2013). Measured by value transferred, it is the world's largest payment system (Committee on Payment and Settlement Systems 2013, Table PS3), surpassing even the largest single-currency systems.

Payments made through CLS occur as transfers on the books of a limited-purpose U.S. bank (CLS Bank) supervised by the Federal Reserve in cooperation with other central banks.[18] CLS has access to the Fed's large-value system (Fedwire) and also to large-value payment systems in all of the currencies it operates in. "Deposits" into (known as *pay-ins*) and "withdrawals" from CLS Bank (*pay-outs*) occur in central bank funds and occur immediately via the appropriate large-value, real-time gross settlement (RTGS) system.[19] Thus, CLS functions as a "daylight bank" with no deposits in its accounts overnight. Payments (account transfers) over CLS can be made by "member" commercial banks in any of its participating currencies, with about 45 percent of CLS payments occurring in U.S. dollars.

Approximately seventy-five banks are members of CLS.[20] Reflecting the immense turnover in the FX markets, daily turnover at CLS is also enormous. Following days of heavy market activity or U.S. legal holidays (when two days of settlements must be compressed into one), the value of payments made through CLS can be breathtaking—the current record daily value is $10.3 trillion on March 19, 2008, in the wake of the Bear Stearns collapse.[21]

4.2 How CLS operates: examples[22]

The special problems of FX settlement, the operation of CLS, and its interaction with traditional large-value payment systems can be illustrated through a series of examples.

---

[18] Actual processing of payments is carried out by a separate U.K. company (CLS Services). Both CLS Bank and CLS Services are owned by a holding company, that is in turn owned by 75 financial institutions worldwide (Committee on Payment and Settlement Systems 2008a).

[19] A notable exception occurs for pay-ins and pay-outs in Canadian dollars, which are sent through a net settlement system (the Large Value Transfer System or LVTS), whose payments are guaranteed by the Bank of Canada. For purposes of the discussion here, these can be regarded as the equivalent of RTGS payments.

[20] CLS also provides indirect settlement services to over 11,000 "third parties," i.e., customers of CLS member banks who must settle through a designated member.

[21] Given these magnitudes, it comes as no surprise that the CLS Bank has been designated a "systemically important financial market utility" by U.S. regulators.

[22] The examples and discussion in this section are taken from Kahn and Roberds (2001), section 2.1.

Example 1. On day *T*, a trader for Bank *A* buys dollars from a trader for Bank *B* in return for pounds. For simplicity, say that the agreed-upon exchange rate is $2/£, and that $2 million is traded for £1 million. Even though this is a "spot" trade of one currency for another, like most financial market trades it is really an exchange of promises to deliver something (in *A*'s case, dollars; in B's case, pounds) in the near future—day *T*+2 for the canonical spot FX trade.[23]

The first difficulty in settling FX trades occurs because there is limited scope (in this initial example, none) for reducing *A*'s and *B*'s settlement exposures through netting: *B* has promised to deliver something (dollar funds) which is (traditionally) only deliverable through the U.S. banking and payment systems, subject to U.S. law, while *A*'s delivery must be routed through U.K. institutions.[24] There exists no natural choice of a "third asset" or numeraire that could serve as the basis for netting. The second difficulty is how to enforce conditionality of settlement without the use of a central counterparty—to span both sides ("legs") of an FX transaction, a traditional central counterparty would need to be able to simultaneously replace trading obligations within the constraints imposed by the national institutions of each leg of the trade. For the present time, such centralization remains an impractical option for most FX trades; see however the discussion below.

The traditional method for settling a foreign exchange trade relies on separate, uncoordinated settlement actions by each party to the trade. To illustrate the traditional process, the first pair of accounts in Table 1 shows the situation after the trade on day *T*.[25]

---

[23] FX trades also commonly occur as forward transactions or FX swaps (a spot combined with a forward). Issues involving settlement of these types of trades are similar to those arising from spot trades.

[24] Again there are exceptions. One is in the case of *non-deliverable forwards*, which are forward trades of a convertible currency (e.g., dollars) against another currency which may be thinly traded or not fully convertible (e.g., yuan). Non-deliverable forwards are typically settled in the convertible currency, as a cash payment in the difference in the contracted value against the spot value of the nonconvertible currency.

[25] For purposes of illustration, Example 1 assumes that each bank directly makes payments over a large-value payment system to settle the hypothetical trade. In fact, banks often effect settlement by instructing a corre-

Bank *A*

Assets	Liabilities
+ $2M due from Bank *B*	+ £1M due to Bank *B*

Bank *B*

Assets	Liabilities
+ £1M due from Bank *A*	+ $2M due to Bank *A*

Table 1a. Settlement of spot FX Trade: traditional system
Bank *A* purchases $2M from Bank *B* for £1M
Shown are banks' positions after trade, before settlement

On day *T*+2, Bank *A* is obligated to send £1 million to *B* over the U.K. large value system (CHAPS) and Bank *B* is obligated to send $2 million to *A* over a dollar payment system (traditionally, CHIPS). Suppose that, due to time zone differences, the sterling transaction happens to be executed first (Table 1b).

Bank *A*

Assets	Liabilities
– £1M in CB funds + $2M due from Bank *B*	

Bank *B*

Assets	Liabilities
+ £1M in CB funds	+ $2M due to Bank *A*

Table 1b. Traditional system after settlement of one leg

In most cases, the dollar funds transfer then occurs, settling the trade, so at the end of the day *T*+2 we have:

spondent to make such payments. Hence the traditional method is referred to as the *correspondent banking method* of settlement.

Bank *A*

Assets	Liabilities
+ $2M CB Funds – £1M CB Funds	

Bank *B*

Assets	Liabilities
– $2M CB Funds + £1M CB Funds	

Table 1c. Traditional system after settlement of second leg

Clearly, the traditional system can lead to problems, given the finality of payments made in each currency.[26] For instance, if Bank *B* is closed down before its funds are sent to Bank *A*, there is the risk that Bank *A* may lose its entire principal in the trade.[27] On the other hand, suppose it is Bank *A* that is shut down early on date *T+2.* Then, in practice, Bank *B* is also likely to suffer a loss even if the shutdown occurs before any settlement takes place, because it can be difficult for either bank to cancel its leg of transaction, should it learn of the failure of its counterparty.[28]

The key precept of CLS is to avoid the possibility of loss of principal by requiring both legs of an FX transaction to settle simultaneously, on the books of a single institution (CLS Bank). While CLS does not formally operate as a central counterparty across currencies, its ability to enforce this conditionality allows it to function in many circumstances as a "virtual central counterparty."

[26] The payments in this example occur over large-value systems where all payments are irrevocable.

[27] In the literature this risk is variously referred to as *principal risk, Herstatt risk, and (cross-country) settlement risk*.

[28] These difficulties are often attributed to the high degree of automation in settlement processes. For example, KfW Bankengruppe, a German state bank, is reported to have sent €300 million to Lehman Brothers as an automated settlement of a swap, on the same day Lehman filed for bankruptcy (Kulish, 2008).

For purposes of illustration, assume that Banks *A* and *B* are both members of CLS, and that no other transactions take place on day *T*. On the morning of day *T*+2, each CLS member is required to make a payment (i.e., a *pay-in*) on its short positions. For the moment we will assume that the payments required are equal to the full value of the trade; more complicated cases are considered later.

Table 2 illustrates the process. Note that *A*'s and *B*'s initial positions following the trade are the same as in the previous case.

Bank *A*

Assets	Liabilities
+ $2M due from Bank *B*	+ £1M due to Bank *B*

Bank *B*

Assets	Liabilities
+ £1M due from Bank *A*	+ $2M due to Bank *A*

CLS Bank

Assets	Liabilities
0	*0*

Table 2a. Settlement of spot FX trade: CLS with full pay-in
Bank *A* Purchases $2M from Bank *B* for £1M
Shown are banks' positions after trade, before settlement

Each bank begins the settlement process by making its pay-in to CLS. These payments are made through RTGS systems in central bank funds—in the example, through Fedwire for the dollar payment and through CHAPS for the sterling payment.

Bank *A*

Assets	Liabilities
+ $2M due from Bank *B* + £1M due from CLS Bank – £1M CB Funds	+ £1M due to Bank *B*

Bank *B*

Assets	Liabilities
+ £1M due from Bank *A* + $2M due from CLS Bank – $2M CB Funds	+ $2 M due to Bank *A*

CLS Bank

Assets	Liabilities (Accounts)		
+ £1M CB Funds		Currency Sub Accts.	
+ $2M CB Funds		£	$
	Bank *A*	£1M	
	Bank *B*		$2M

Table 2b. CLS system after full pay-in

Once CLS has both currencies available to it, settlement is effected through a paired set of payments on the books of CLS, as is shown in Table 2c.[29] These payments occur automatically once there are sufficient funds in each bank's account. Note that settlement is on a gross basis; each bank pays and receives the full amount of the funds due in the trade, in the form of balances on the books of the CLS Bank.

[29] Formally, payment over CLS does not constitute legal settlement of FX trades but of "the payment instructions arising from the trades" (Committee on Payment and Settlement Systems 2008a, 24, fn 31). For our purposes, the distinction is inessential.

Bank *A*

Assets	Liabilities
+ $2M due from CLS Bank – £1M CB Funds	

Bank *B*

Assets	Liabilities
+ £1M due from CLS Bank – $2M CB Funds	

CLS Bank

Assets	Liabilities (Accounts)		
+ £1M CB Funds		Currency Sub Accts.	
+ $2M CB Funds		£	$
	Bank *A*		$2M
	Bank *B*	£1M	

Table 2c. CLS system after settlement

After that, the currencies can be sent out to the banks via the same RTGS systems that were used for the pay-ins:

Bank *A*

Assets	Liabilities
+ $2M CB Funds – £1M CB Funds	

Bank *B*

Assets	Liabilities
+ £1M CB Funds – $2M CB Funds	

CLS Bank

Assets	Liabilities
0	*0*

Table 2d. CLS system after pay-out

Under CLS, final settlement of each side of a transaction is simultaneous and mutually conditional (in payments jargon, this feature is known as *payment versus payment* or PVP, similar to *delivery versus payment* or DVP for domestic securities transactions). As this example shows, with CLS there is never a point at which one leg is settled and the other is not settled. Under the traditional arrangement there is an instant where one bank (Bank *A* in Table 1b) has paid out funds to its counterparty but not received funds in return. Were Bank B to fail at this moment then Bank *A,* as its creditor, would be vulnerable. By contrast, at no point in the process in Table 2 is either bank a net creditor of the other. Under CLS, if Bank *B* fails before settlement, the transaction does not go through, and the funds paid in by Bank *A* are returned to Bank *A.* If Bank *B* fails after settlement, Bank *A* is unaffected.

Of course after settlement, Bank *A* is now a creditor of CLS until CLS sends it the payments to Bank *A.* CLS is an improvement over traditional arrangements, because CLS Bank is a better credit risk than any individual bank. In this simple example, because CLS Bank is never the creditor of any bank, it is invulnerable to failures of other banks. The finality of payments on RTGS systems is key to this arrangement. Because the CLS Bank's assets are simp-

ly “good funds,” not “due froms,” they won’t disappear if the bank that paid them in goes bankrupt.

Example 1 is an extreme case: both sides pay in full before settlement. In practice, CLS allows members to overdraft their accounts, so that settlement may occur before all net funds have been paid in.[30] As a result, a (very modest) level of risk creeps back into the arrangement. The next example considers a case where settlement takes place after only a small initial pay-in. Table 3 below illustrates the settlement process for this example.

Example 2. As before, but suppose that initially Bank *B* pays in $200,000 or 10% of its due-to position in its short currency, and Bank *A* pays in a corresponding amount: £100,000. As before, settlement occurs by transferring the required balances between the sub-accounts of the two banks on the books of CLS: £1M from Bank *A*’s sterling sub-account to Bank *B*’s sterling sub-account and $2M from Bank *B*’s dollar sub-account to bank *A*’s dollar sub-account. Now, however, these transactions leave overdrafts in a sub-account for each of the banks. Once sufficient pay-ins are made, the situation is the same as in Example 1, and pay-out can proceed safely. But until pay-in is completed, the system is vulnerable to a failure by either of the banks. For example, if Bank *B* fails before completing its pay-in, CLS Bank will owe Bank *A* $1 million but will only have $100,000 in good funds.

[30] CLS however sets a minimum pay-in schedule for each member on a daily basis.

Bank *A*

Assets	Liabilities
+ $2M due from Bank *B*	+ £1M due to Bank *B*

Bank *B*

Assets	Liabilities
+ £1M due from Bank *A*	+ $2M due to Bank *A*

CLS Bank

Assets	Liabilities
0	*0*

Table 3a. Settlement of spot FX trade: CLS with delayed pay-In
Bank *A* purchases $2M from Bank *B* for £1M
Positions shown are after trade, before settlement

Assets	Liabilities
+ $2M due from Bank *B* + £0.1M due from CLS Bank – £0.1M CB Funds	+ £1M due to Bank *B*

Bank *B*

Assets	Liabilities
+ £1M due from Bank *A* + $0.2M due from CLS Bank – $0.2M CB Funds	+ $2M due to Bank *A*

CLS Bank

Assets	Liabilities (Accounts)	
+ £0.1M CB Funds + $0.2 M CB Funds	Currency Sub Accts.	
	£	$
	Bank *A* £0.1M	
	Bank *B*	$0.2 M

Table 3b. After initial pay-in by Bank A

Bank *A*

Assets	Liabilities
+ $2M due from CLS Bank – £0.1M CB Funds	+ £0.9M Overdraft at CLS

Bank *B*

Assets	Liabilities
+ £1M due from CLS Bank – $0.2M CB Funds	+ $1.8 M Overdraft at CLS

CLS Bank

Assets	Liabilities (Accounts)		
+ £0.1M CB Funds + $0.2M CB Funds		Currency Sub Accts.	
		£	$
	Bank *A*	–£0.9M	$2M
	Bank *B*	£1M	– $1.8 M

Table 3c. After settlement

Table 3d. After final pay in (Identical to Table 2c)

Table 3e. After pay-out (Identical to Table 2d)

Although CLS permits member banks to have overdrafts during the settlement process, the overdrafts are subject to limits. A transaction is not settled if it causes a member to exceed its position limits; instead both legs of the transaction are held in a queue until sufficient funds flow into the bank's account. The overdraft limits include limits on each sub account, as well as a separate limit on the sum of overdrafts.[31] Most importantly, a member's net position *across all currencies* is required to be positive at all times. Again, CLS Bank is never in the position of being an overall creditor to any member bank. Thus failure of Bank *B* does not adversely affect the value of the CLS Bank.

[31] The limit on the sum of the overdrafts is called the member's *aggregate short position limit.* It is adjusted by CLS according to factors such as capital and credit rating of the member (Committee on Payment and Settlement Systems 2008a, 79-80).

In order to handle the possibility of a failure by a bank with an overdraft, the CLS Bank has arranged lines of credit in each of its currencies with a set of "liquidity providers."[32] Since CLS essentially carries no credit risk, it can obtain these credit lines at extremely small costs. It is clear why the liquidity providers can trust CLS Bank; it is less clear why the CLS Bank should be satisfied with the reliability of its liquidity providers—who turn out to be owners of the CLS Bank, that is, the member banks themselves. Then might the protection offered by them be illusory? There are two counterarguments: first it is the group of liquidity providers as a whole that provides protection to the CLS bank against failure of any individual member. Second, the limits on overdrafts under CLS, while explicitly protecting existing liquidity providers, also serve to convince any potential additional liquidity providers—conceivably including, in extreme situations, central banks—of the ultimate safety of their liquidity infusions.

In the absence of exchange rate fluctuations, settlement could begin before the pay-in of any funds, without violating the principle that a bank's net position at CLS must not be negative. When exchange rates fluctuate, the "out-of-the-money party" (at least) must make some pay-in before settlement can begin.[33]

Example 2, continued. Suppose that on day T+1 the value of the pound falls, so that by the close of trading on day T+1, £1 is now worth only \$1.80, so that Bank *B*'s position vis-à-vis Bank *A* is "out of the money." CLS then calculates the day T+2 settlement schedule as follows: Bank *B* is short dollars so has a minimum pay-in obligation of \$200,000 or 10% of its due-to position in its short currency, before processing of its transactions can begin.[34]

[32] CLS generally has contracts with at least three liquidity providers in each currency.

[33] Thus the net positive balance requirement plays much the same role as margin requirements under "marking-to-market" in a futures clearing arrangement. See e.g., Baer, France, and Moser, (2004), or Moser (1998).

[34] In practice, CLS levies an additional haircut to provide a cushion against exchange rate volatility during day T+2.

Thus far, the examples have dealt with a single payment. In fact participants in FX markets make large numbers of exchanges during the day, repeatedly swapping currencies back and forth in offsetting or near-offsetting trades.

Example 3. Suppose that Bank *A* buys $4 million from Bank *B* for £2 million during the first trade of day *T*, and then buys £1 million from Bank *B* for $2 million during the second trade of the same day, with all trades at $2/ £. As before, assume that the dollar rises to £1 = $1.80 by the close of trading on day *T*+1, so that each bank's initial pay-in requirement would be the same as in example 2. That is, at the beginning of day *T*+2, Bank *B* is net short $2 million and long £1 million, so once again *B* would need to pay in $200,000.

When settlements occur depends on the size of the two banks' permitted overdrafts. For simplicity we will assume that overdrafts are sufficient to handle each trade. Nonetheless, a trade cannot settle until the pay-ins are adequate to ensure that each bank have a net positive post-trade balance. Since bank *B* is out of the money $400,000 on the first trade, that trade will not settle until bank *B* puts a further $200,000 in its account. Once it does, the first trade will settle. Although *A* is out of the money on the second trade, the settlement of the first trade leaves the net position of *A* sufficiently positive to enable the second trade to settle as well.

Although the trades are settled, the CLS Bank still lacks the funds to make a payout. These must await the pay-in of additional funds by each of the banks. As those funds appear, pay-outs are made on settled trades subject to two restrictions: 1) the CLS Bank can never overdraw its account with any RTGS system, and 2) all settlement banks' accounts with CLS must remain net positive. Pay-ins, settlement, and pay-outs continue on an ongoing basis until all transactions have been settled and all funds paid out.[35]

[35] The exact choice of which transactions to pay out first is made according to a proprietary algorithm. The algorithm accords preference to members and currencies with the highest balances and to currencies with the earliest large-value payment system closing times (Committee on Payment and Settlement Systems 2008a, 78).

4.3 Liquidity saving and In-Out Swaps

For many of its participants a major advantage of the CLS system is the opportunity it provides to economize on the use of currency through the "liquidity recycling" arrangement described in example 3 above. For the purpose of settlement the CLS arrangement is *not* a netting arrangement. Each trade settles or fails separately: Given a pair of bilateral trades between two banks, it would be possible for one to settle and the other not to, due for example to a subsequent failure of a bank to make a pay-in. On the settled trade, payouts become the responsibility of CLS Bank. On the unsettled trades, each bank is returned its initial pay-ins. Nonetheless, CLS shares one important feature with traditional netting arrangements: it economizes on the use of central bank funds. In our simple example, each bank only need pay in its net position in the short currency for CLS to be able to complete the payment process. With stricter caps on overdraft positions a greater pay-in may be required, but as a bank engages in larger numbers of transactions the difference becomes small.

This "quasi-netting" property of CLS settlement generates liquidity savings that are comparable to the savings available through multilateral net settlement. According to the CLS website, quasi-netting reduces pay-in amounts to about four percent of the gross amounts due. But (.04) × $5 trillion is still a lot of money, even by the rarefied standards of today's large-value payment systems. CLS's need for liquidity is exacerbated by its need for immediacy: to enable simultaneous worldwide settlement, CLS must begin processing payments very early in the American and European business days, when traditionally little liquidity is available except through central banks.[36]

CLS's liquidity demand could, in principle, be entirely met by borrowing from central banks, but CLS member banks have been reluctant to tie up their available intraday credit capacity in this fashion. Instead, as a way of reducing the liquidity costs of CLS pay-ins, they have de-

[36] Pay-ins to CLS begin at 7 a.m. Central European Time. Settlement begins at the same time and is normally complete by 9 a.m. CET, but pay-outs (and additional pay-ins) may continue until 10 a.m. CET for Asian currencies, and noon CET for all other currencies.

veloped a private intraday lending mechanism known as the *in-out swap.* In-out swaps are coordinated through CLS but are technically side agreements that are outside of the CLS system. An in-out swap consists of a pair of transactions that occur on the same day. In the first transaction, a CLS member exchanges, within CLS, a position in a currency in which it is long against a currency in which it is short, thereby reducing both its pay-ins and pay-outs. The second transaction happens later the same day and occurs outside CLS, and reverses the first transaction at exactly the same exchange rates. As with other intraday credit mechanisms, these intraday swaps are not priced and traded; CLS identifies potential swaps the night before and members are free to agree to exchanges "at par" or not.

By using in-out swaps, CLS settlement members have been able to reduce the liquidity required for their pay-ins to less than two percent of gross amounts due (Committee on Payment and Settlement Systems 2008a), i.e., less than $100 billion equivalent across all currencies on an "average" CLS day.[37] As usual there is no free lunch: since the "outside" transaction in an in-out swap is settled through the traditional "correspondent banking" method of settling FX transactions, the use of in-out swaps represents a partial retreat from the conditionality guarantee of the CLS system. Discussions of this issue usually point out that the residual amount of principal or Herstatt risk that has been reintroduced by the use of in-out swaps is small relative to the risk present before the introduction of CLS.

4.4 Policy issues for central banks

The foregoing discussion shows how payments made through CLS can substitute for payments in central bank money, and provide protection against principal risk in situations where traditional forms of FX settlement could not. The design of CLS, while robust, cannot protect against all types of risks in FX trades in all situations. In particular, CLS cannot guarantee the liquidity of its participants. A CLS member might, for example, fail to pay in its obligation, in which case CLS deletes that member's trades from its system. This protects

[37] For single-currency, large-value payment systems a common ratio of net to gross payments is approximately one percent (Bech and Hobijn 2007), which would represent a lower limit on liquidity needed for CLS settlement. CLS does not quite attain this limit but comes close.

the principal of the remaining members but may subject them to liquidity pressures due to unexpected shifts in their pay-in requirements. Similarly, a failure of multiple liquidity providers in a given currency could lead to widespread stresses. Thanks to the rule that CLS is never in a negative net position, such an event would not endanger the solvency of CLS, but its ability to make pay-outs in the affected currency could be impaired. In such cases, CLS rules allow for the CLS Bank to complete pay-outs in currencies where sufficient liquidity is available. This again would preserve principal but possibly subject the remaining members to unexpected liquidity demands.

The examples presented earlier should also make clear that any CLS-induced liquidity strains would not necessarily be confined to a single currency. A failure by one member to pay in say, Euro to its CLS account, could lead to a short of liquidity and cause disruptions to large-value payment systems in other currencies. Defenders of CLS have pointed out that cross-currency linkages existed before but were only less apparent, and, because they did not control principal risk, were potentially even more disruptive. However, the ultimate allocation of residual risks, and the extent to which these are borne by central banks, is yet to be resolved.

To date, doubts about the integrity of CLS settlement have remained in the realm of the hypothetical. Notably, CLS was able to continue normal settlement processes in the wake of the market disruptions of 2007 and 2008. A watershed event was the September 2008 failure of Lehman Brothers. Lehman was a "user member" of CLS that relied on another CLS member (Citigroup) for settlement services. The decision by Citigroup to continue to settle the failed member's trades enabled CLS settlement to proceed without disruption. However, use of in-out swaps is reported to have contracted in wake of the Lehman bankruptcy, leading to some reduction in liquidity savings (Foreign Exchange Contact Group and Operations Managers Group 2009).

CLS' ability to withstand the shocks experienced during the recent crisis appears to have blunted movement toward additional centralization of FX clearing. Notably, a recent ruling by U.S. regulators (United States Treasury 2012) has granted FX markets a specific exemp-

tion from the clearing requirements of the 2010 Dodd-Frank Act. The ruling makes frequent mention of the efficacy of CLS in controlling settlement risk.

The main business of CLS is settling foreign exchange transactions, but it has branched out into other activities. In early 2008 it launched a service (in cooperation with DTCC[38]) for settling credit derivatives trades. By virtue of CLS' connections to multiple large-value payment systems, there is no technological barrier to using it to settle other types of trades as well. Another unresolved policy issue is to what extent future expansions of CLS would be consistent with its original purpose of managing risks in FX markets.

5. Conclusion

From their beginnings, central banks have had a role in payments. This role has rarely been static, however, and as central banks have innovated, these innovations have been matched, and indeed in many cases outpaced, by the private sector. The result has been a steady if not always monotone progression toward lower costs and reduced risks in payments. We have considered three examples of this process.

The first example described the payments role of the Bank of Amsterdam, the most prominent of the Early Modern "exchange banks"—account-based public banks whose principal function was settlement of a form of private payments (bills of exchange) prevalent during that era. In 1683, the Bank of Amsterdam enacted a reform which provided its users cheap access to liquidity, and so was able to take on a dominant payments role within eighteenth-century European commerce. Ultimately the Bank failed, however, because it could not successfully reconcile its payments role with demands on it from fiscal authorities.

Our second example described the payments role of an ultimately more successful institution, the Bank of England. Like the exchange banks, the Bank of England offered accounts, but more important to its operations were the bearer notes that it issued on an unprece-

[38] *D*epository *T*rust and *C*learing *C*orporation, which owns several major U.S. financial market utilities.

dented scale. Lacking the legal and financial resources to compete as note issuers, private banks responded by developing check payments into a viable alternative to notes. However, the private banks remained dependent on the Bank of England for settlement services, and especially for access to liquidity during financial crises. The nineteenth-century U.S. banking system sought to imitate the British success with checks, but a lack of a strong central bank made the highly fragmented American system susceptible to frequent crises, and therefore less attractive to international participants.

Modern RTGS systems retain aspects of both of these earlier systems, and remain the backbone of payments in most countries. Yet our third example shows how a private-sector payment system, CLS, has been able to take payments beyond the confines of any single-currency system. Through an innovative design, CLS has reduced the chances that FX market participants will suffer a loss of principal in a trade. For FX transactions, CLS is now the anchor; central banks play a vital, but secondary role in this design. But CLS has also helped to increase the interconnectedness of the world's large-value payment systems. The end result may be only to extend central banks' responsibilities for the integrity of payments.

REFERENCES

Arango, Carlos, and Angelika Welte. 2012. "The Bank of Canada's 2009 Methods-of-Payment Survey: Methodology and Key Results." Discussion Paper 2012-6, Bank of Canada.

Baer, H., France, V.G., and Moser, J.T., (2004). "Opportunity Cost and Prudentiality: an Analysis of Collateral Decisions in Bilateral and Multilateral Settings," *Research in Finance* 21: 201-227.

Berger, Allen, Dianna Hancock and Jeffrey Marquardt. 1996. "A Framework for Analyzing Efficiency, Risks, Costs, and Innovations in the Payments System." *Journal of Money Credit and Banking* 28: 696-732.

Bech, Morten L. and Bart Hobijn. 2007. "Technology Diffusion within Central Banking: the Case of Real-Time Gross Settlement." *International Journal of Central Banking* (September): 147-181.

Bech, Morten and Jhuvesh Sobrun. 2013. "FX market trends before, between, and beyond Triennial Surveys." *BIS Quarterly Review* (December), 45-54.

Bidermann, Hermann I., 1859. *Die Wiener Stadt-Bank.* Archiv für Kunde österreichischer Geschichtsquellen, Vienna.

Bignon, Vincent, Marc Flandreau, and Stefano Ugolini. 2012. "Bagehot for Beginners: the Making of Lender-of-Last-Resort Operations in the Mid-Nineteenth Century." *Economic History Review* 65(2): 580-608.

Bodenhorn, Howard. 1992. "Capital mobility and financial integration in antebellum America." *Journal of Economic History* 52, no. 3:585–610.

Bodenhorn, Howard. 2001. "Antebellum Banking in the United States". *EH.Net Encyclopedia*, edited by Robert Whaples. September 5, 2001. URL http://eh.net/encyclopedia/article/bodenhorn.banking.antebellum

Börner, Lars and John William Hatfield. 2012. "The Design of Debt Clearing Markets: Clearinghouse Mechanisms in Pre-Industrial Europe."

Broz, J. Lawrence. 1999. Origins of the Federal Reserve System: International Incentives and the Domestic Free-Rider Problem." *International Organization* 53(1): 39-70.

Broz, J. Lawrence, and Richard S. Grossman. 2004. "Paying for the Privilege: the Political Economy of Bank of England Charters, 1694-1844." *Explorations in Economic History* 41: 48-72.

Brunnermeier, Marcus K. and Lasse H. Pedersen, "Market Liquidity and Funding Liquidity." *Review of Financial Studies,* 2009.

Cannon, James Graham. 1910. *Clearing Houses.* Vol. 6 of the Report by the National Monetary Commission to the U.S. Senate, 61st Cong., 2nd sess., Doc. 491, Washington, D.C., Government Printing Office.

Carlos, Ann M. and Larry Neal. 2011. "Amsterdam and London as financial center in the eighteenth century." *Financial History Review* 18(1): 21-46.

Carter, Susan B., Scott Sigmund Gartner, Michael R. Haines, and Alan L. Olmstead (eds). 2006. *Historical Statistics of the United States.* Cambridge University Press, Cambridge.

Cavalcanti, Ricardo De O., and Neil Wallace. 1999. "Inside and Outside Money as Alternative Media of Exchange" *Journal of Money, Credit and Banking*, Vol. 31, No. 3, pp. 443-457.

Cavalcanti, Ricardo de O., Andres Erosa, and Ted Temzelides, 1999. "Private Money and Reserve Management in a Random-Matching Model," *Journal of Political Economy* 107(5): 929-945.

Clapham, John, 1944. *The Bank of England, a history*. Cambridge University Press, Cambridge.

Committee on Payment and Settlement Systems. 1980. *Payment Systems in Eleven Developed Countries.* Bank for International Settlements, Basel.

Committee on Payment and Settlement Systems. 1996. "Settlement risk in foreign exchange transactions." Basel, Bank for International Settlements.

Committee on Payment and Settlement Systems. 1998. "Reducing foreign exchange settlement risk: A progress report." Basel, Bank for International Settlements.

Committee on Payment and Settlement Systems. 2002. *Statistics on payment and settlement systems in selected countries: Figures for 2000*. Bank for International Settlements, Basel.

Committee on Payment and Settlement Systems. 2008a. *Progress in reducing foreign exchange settlement risk.* Basel, Bank for International Settlements.

Committee on Payment and Settlement Systems. 2013. *Statistics on payment, clearing, and settlement systems in the CPSS countries: Figures for 2012.* Bank for International Settlements, Basel.

Dehing, Pit. 2012. *Geld in Amsterdam: Wisselbank en wisselkoersen, 1650-1725*. Uitgeverij Veloren, Hilversum.

Denzel, Markus A., 2012. *Der Nürnberger Banco Publico, seine Kaufleute und ihr Zahlungsverkehr.* Franz Steiner Verlag, Stuttgart.

De Vries, Jan, and Ad Van der Woude. 1997. *The first modern economy: success, failure and perseverance of the Dutch economy, 1500-1815.* Cambridge University Press, Cambridge.

Eichengreen, Barry, and Marc Flandreau, 2012. "The Federal Reserve, the Bank of England, and the Rise of the Dollar as an International Currency, 1914-1939." *Open Economies Review* 23(1): 57-87.

Flandreau, Marc, and Clemens Jobst. 2005. "The Ties that Divide: A Network Analysis of the International Monetary System, 1890–1910." *Journal of Economic History* 65: 977-1006.

Flandreau, Marc, Christophe Galimard, Clemens Jobst and Pilar Nogués-Marco. 2009. "Monetary Geography before the Industrial Revolution." *Cambridge Journal of Regions, Economy and Society* 2:149-171.

Foreign Exchange Contact Group and Operations Managers Group. 2009. *Report on Operational Lessons from the Demise of Lehman Brothers in Autumn 2008.* Frankfurt, European Central Bank.

Foster, Kevin, Erik Meijer, Scott Schuh, and Michael A. Zabek, 2011, "The 2009 Survey of Consumer Payment Choice." Public Policy Discussion Paper 11-1, Federal Reserve Bank of Boston .

Geanakoplos, John. 2009. “The Leverage Cycle” NBER Macroeconomics Annual Vol. 24, No. 1.

Giannini Curzio, 2011. *The Age Of Central Banks*, Edward Elgar.

Gilbert, R. Alton. 2000. The advent of the Federal Reserve and the efficiency of the payments system: The collection of checks 1915–1930. *Explorations in Economic History* 37, no. 2:121–48.

Gorton, Gary. 1996. “Reputation Formation in Early Bank Note Markets.” Journal of Political Economy 104(2): 346-397.

Grossman, Richard. 2003. "US Banking History, Civil War to World War II". *EH.Net Encyclopedia*, edited by Robert Whaples. March 17, 2003. URL http://eh.net/encyclopedia/article/grossman.banking.history.us.civil.war.wwii

Harris, Ron. 2004. “The Bubble Act: It’s Passage and Its Effects on Business Organization.” *Journal of Economic History* 54: 610-27.

Heckscher, Eli F. 1934. “The Bank of Sweden in its connection with the Bank of Amsterdam. ” In: J.G. van Dillen (Ed.), *History of the principal public banks. Martinus Nijhoff,* The Hague, pp. 161-199.

Hills, Sally, Ryland Thomas, and Nicholas Dimsdale. 2010. “The UK recession in context – what do three centuries of data tell us?” *Bank of England Quarterly Bulletin* 50(4), 277-291.

Holmström, Bengt and Jean Tirole. 2011. *Inside and Outside Liquidity*, MIT Press.

James, John. 2012a. “English Banking and Payments before 1826.” *Research in Economic History* 28: 117-49.

James, John. 2012b. “Panics, Payments Disruptions and the Bank of England before 1826.” *Financial History Review* 19(3): 289-309.

James, John and David F. Weiman. 2010. “From Drafts to Checks: the Evolution of Correspondent Banking Networks and the Formation of the Modern U.S. Payments System.” *Journal of Money, Credit and Banking* 42: 237-265.

James, John and David F. Weiman. 2011. “The National Bank Acts and the Transformation of New York City Banking During the Civil War Era.” *Journal of Economic History* 71(2): 338-62.

James, John, James McAndrews, and David F. Weiman. 2013. “Wall Street and Main Street: the Macroeconomic Consequences of New York Bank Suspension, 1866-1914.” *Cliometrica* 7: 99-130.

Jaremski, Matthew. 2011. “Bank-Specific Default Risk in the Pricing of Bank Note Discounts.” *Journal of Economic History* 71(4): 950-975.

Kahn, Charles M., James McAndrews, and William Roberds. 2005. “Money is Privacy.” *International Economic Review* 46(2): 377-99.

Kahn, Charles M. and William Roberds. 2001. “The CLS Bank: A Solution to the Problem of International Payments Settlement?” *Carnegie-Rochester Conference Series on Public Policy,* 191-226.

Kahn, Charles M. and William Roberds. “Why Pay? An Introduction to Payments Economics.” *Journal of Financial Intermediation* 2009 18(1) 1-23.

Kahn, Charles M. 2009. "Collateral Policy in a World of Round-The-Clock Payment" working paper, University of Illinois Department of Finance University of Illinois College of Business working paper 10-0100.

Kahn, Charles M. 2013. "Private Payment Systems, Collateral, and Interest Rates." *Annals of Finance*, Volume 9, Issue 1, Page 83-114.

Kahn, Charles M. and José Manuel Liñares-Zegarra. 2012. "Identity Theft and Consumer Payment Choice: Does Security Really Matter?" working paper, University of Illinois, Urbana-Champaign.

Klee, E., 2008. **"**How people pay: Evidence from grocery store data**"** *Journal of Monetary Economics*.

Kleer, Richard. 2012. "The Folly of Particulars: the Political Economy of the South Sea Bubble." *Financial History Review* 19(2): 175-197.

Knodell, Jane. 2003. "Duty and Profit in the Second Bank of the United States Exchange Operation." *Financial History Review* 10, no. 1:5–30.

Kocherlakota, 2001, "Risky Collateral and Deposit Insurance." Advances in Macroeconomics 1, art 1 1.

Kosse, Anneke. 2012. "Do Newspaper Articles on Card Fraud Affect Debit Card Usage?" research paper, Cash and Payment Systems Division De Nederlandsche Bank, the Netherlands Version: March 2012.

Kulish, Nicolas. 2008. "Uproar over German Bank's Payout to Lehman." *New York Times*, September 18.

Lagos, Ricardo and Randall Wright. 2005. A unified framework for monetary theory and policy analysis. *Journal of Political Economy* 113, 463-484.

Lagos, Ricardo. 2006. "Inside and Outside Money.," Federal Reserve Bank of Minneapolis Research Department Staff Report 374, May 2006.

Leinonen, Harry (ed.). 2008. *Evolving payment habits* Proceedings of the Bank of Finland Payment Habits Seminar, 2008.

Luzatto, G. 1934. "Les Banques publiques de Venise (Siècles XVI-XVIII)." In: J.G. van Dillen (Ed.), *History of the principal public banks.* Martinus Nijhoff, The Hague, pp. 39-78.

Martin, Antoine and James McAndrews, "An Economic Analysis of Liquidity-Saving Mechanisms." *Federal Reserve Bank of New York Economic Policy Review*, September 2008 p 25-39.

Matthews, Philip W. 1921. *The Bankers' Clearing House: What it is and what it does.* Sir Isaac Pitman & Sons, London.

McMahon, Dinny. 2014. "With Yuan Scarce, Firms Get Stuck with IOUs." *Wall Street Journal*, April 4.

Mees, Willem C., 1838. *Proeve eener geschiedenis van het bankwezen in Nederland geduerende den tijd der Republiek.* W. Messcuert, Rotterdam.

Mitchener, Kris and Gary Richardson. 2013. "Shadowy Banks and Financial Contagion during the Great Depression: a Retrospective on Friedman and Schwarz." *American Economic Review* 103(3): 73-78.

Moser, James T. 1998. "Contracting innovations and the evolution of clearing and settlement methods at futures exchanges." Working paper, Federal Reserve Bank of Chicago.

Neal, Larry. 1998. "The Financial Crisis of 1825 and the Restructuring of the British Financial System." *Federal Reserve Bank of St. Louis Review* 80(3): 53-76.

Neal, Larry and Stephen Quinn. 2001. "Networks of Information, Markets, and Institutions in the Rise of London as a Financial Center, 1660-1720," *Financial History Review* 8: 7-26.

Nogues-Marco, Pilar. 2013. "Competing Bimetallic Ratios: Amsterdam, London, and Bullion Arbitrage in Mid-Eighteenth Century." Journal of Economic History 73(2): 445-76.

Norman, Ben, Rachel Shaw, and George Speight. 2011. "The history of interbank settlement arrangements: exploring central banks' role in the payment system" Working Paper No. 412, Bank of England.

Nosal, Ed, and Guillaume Rocheteau, 2011. *Money, Payments, And Liquidity*. MIT Press.

Polak, Menno. 1998. *Historiografie en Economie van de "Muntchaos."* NEHA, Amsterdam.

Quinn, Stephen. 1994. "Tallies or Reserves? Sir Francis Child's Balance Between Capital Reserves and Extending Credit to the Crown, 1685-1695," *Business and Economic History* 23 (1): 39-51.

Quinn, Stephen. 1997. "Goldsmith-Banking: Mutual Acceptances and Inter-Banker Clearing in Restoration London," *Explorations in Economic History* 34: 411-32.

Quinn, Stephen and William Roberds. 2003. "Are On-Line Currencies Virtual Banknotes?" *Federal Reserve Bank of Atlanta Economic Review* (2003): 1-15.

Quinn, Stephen and William Roberds. 2009. "An Economic Explanation of the Early Bank of Amsterdam, Debasement, Bills of Exchange and the Emergence of the First Central Bank." In *The Evolution of Financial Institutions from the Seventeenth to the Twentieth-First Century* 2009, ed. Jeremy Atack and Larry Neal, 32-70. Cambridge University Press.

Quinn, Stephen and William Roberds. 2012. "Responding to a Shadow Banking Crisis: The Lessons of 1763." Federal Reserve Bank of Atlanta Working Paper 2012-8.

Quinn, Stephen and William Roberds. 2014a. "Death of a Reserve Currency." Federal Reserve Bank of Atlanta Working Paper.

Quinn, Stephen and William Roberds. 2014b. "How Amsterdam got Fiat Money." *Journal of Monetary Economics.*

Raudnitz, Josef, 1917. *Das Österreichische Staatspapiergeld und die Priviligierte Nationalbank: erster Theil 1762 bis 1820.* Kaiserliche und königliche Staatsdruckerei, Vienna.

Richardson, Gary. 2007. "The Check is in the Mail: Correspondent Clearing and the Collapse of the Banking System, 1930 to 1933." *Journal of Economic History* 67(3): 643-71.

Riesser, Jacob. 1911. *The German great banks and their concentration in connection with the economic development of Germany.* Vol. 14 of the Report by the National Monetary Commission to the U.S. Senate, 61st Cong., 2nd sess., Doc. 593. Government Printing Office, Washington, D.C.

Ritschl, Albrecht and Mark Spoerer. 1997. "Das Bruttosozialprodukt in Deutschland nach den amtlichen Volkseinkommens- un Sozialproduktsstatistiken 1901-1995." Accessed at www.gesis.org.

Roberds, William and François. R Velde. 2014. "Early public banks." Federal Reserve Bank of Chicago Working Paper No. 2014-03.

Rockoff, Hugh. 2000. "Banking and Finance, 1789-1914." *Cambridge Economic History of the United States*, Vol. II: 643-684.

Santarosa, Veronica. 2010. "Financing Long-Distance Trade Without Banks: The Join-Liability Rule and Bills of Exchange in 18th-Century France." Yale Working Paper

Schuh, S. and J Stavins. 2010. "Why are (some) consumers (finally) writing fewer checks? The role of payment characteristics*" Journal of Banking & Finance* 34: 1745-1758.

Sieveking, Heinrich, 1934a. "Das Bankwesen in Genua und die Bank von San Giorgio." In: J.G. van Dillen (Ed.), History of the principal public banks. Martinus Nijhoff, The Hague, pp. 15-38.

Sieveking, Heinrich, 1934b. "Die Hamburger Bank." In: J.G. van Dillen (Ed.), *History of the principal public banks.* Martinus Nijhoff, The Hague, pp. 125-160.

Siklos, Pierre. 1995. *Great Inflations of the 20th Century.* Edward Elgar, Aldershot.

Stähler, Paul. 1909. *Der Giroverkehr: seine Entwicklung und international Ausgestaltung.* Leipzig, A. Deichert.

Stasavage, David. 2008. *Public Debt and the Birth of the Democratic State*. Cambridge University Press.

Stigler, Stephen M. 1999. *Statistics on the table: the history of statistical concepts and methods.* Harvard University Press, Cambridge, Mass.

Temin, Peter and Hans-Joachim Voth. 2013. *Prometheus Shackled.* Oxford University Press, New York.

Uittenbogaard, Roland. 2009. "Lending by the Bank of Amsterdam (1609-1802)." In *The Bank of Amsterdam*, ed. Marius van Nieuwkerk, 120-131. Sonsbeek, Amsterdam.

United States Department of the Treasury. 2012. "Determination of Foreign Exchange Swaps and Foreign Exchange Forwards Under the Commodity Exchange Act." *Federal Register* 77(224), 69694-69705.

Usher, Abbott P. 1943. *The early history of deposit banking in Mediterranean Europe.* Harvard University Press, Cambridge.

Velde, François. R. 2007. "John Law's system." *American Economic Review* 97, 276-279.

Velde, François. R. 2009. "The case of the Missing Public Bank: Early Modern France." Draft presented at the XVth World Economic History Congress, Utrecht.

Van Dillen, Johannes G., 1934. "The Bank of Amsterdam." In: J.G. van Dillen (Ed.), *History of the principal public banks.* Martinus Nijhoff, The Hague, pp. 79-124.

Wallace Neil. 1983. "A Legal Restrictions Theory of the Demand for 'Money' and the Role of Monetary Policy," Federal Reserve Bank of Minneapolis Quarterly Review Winter 19, Vol. 7 no. 1, pp. 1-7.

Weber, Warren. 2003. “Interbank Payments Relationships in the Antebellum United States: Evidence From Pennsylvania.” *Journal of Monetary Economics* 50(2): 455-474.

Wright, Robert. 2001. "Origins of Commercial Banking in the United States, 1781-1830". *EH.Net Encyclopedia*, edited by Robert Whaples. August 15, 2001. URL http://eh.net/encyclopedia/article/wright.banking.commercial.origins