Payment systems – history and challenges

Paola Boel*

The author is a Senior Economist in the Research Division of the Riksbank

Payment systems are currently undergoing important changes, mostly because of technological innovations. Such changes include a declining role for cash and a growing prominence of fast payment solutions. In this article, I discuss these developments and the challenges they create for central banks. I do so taking into consideration the historical evolution of payment systems and the insights derived from the literature on the economics of payments.

1 Introduction

A *payment* occurs when one party, the payer, transfers an asset to another party, the payee, for the purpose of discharging a debt incurred by the payer. Or, a payment may consist of the payer's instruction to a third party to make such a transfer.[1] While in principle a payment may be made with any asset, in practice virtually all modern payments involve transfers of debt claims on either a central bank ('outside money' in the form of both currency and deposits) or private banks ('inside money', today almost always in the form of deposits).[2]

A *payment system*, in turn, consists of a set of technologies, laws, and contracts that allow payments to occur and determine when a payment constitutes settlement. Payment systems include currency, checks, credit and debit cards, electronic funds transfers, internet banking, and so on. Developed economies depend critically on the efficient functioning of such systems. On the one hand, by offering debtors low-cost and trustworthy means of settling their debts, payment systems stimulate the use of credit, and thus economic activity more generally. On the other hand, unsafe and inefficient payment systems may hamper the efficient transfer of funds among individuals and economic actors (see Humphrey et al., 2006).

But how did central banks come to assume a position at the heart of payment systems? In the first section of this article, I follow Manning et al. (2009) and Norman et al. (2011) to provide an overview of the historical evolution of payment systems, beginning with the early emergence of money and banking. From these foundations, I briefly discuss the subsequent development of banks, and of payment and settlement systems. I then describe the evolution of central banks in the context of their settlement function. I also touch upon the historical reasons for why central banks have been granted the sole right to issue cash.[3] In the second section of the article, I provide a non-technical review of the segment of the payments economics literature that studies both the optimal provision of payment systems and the government's role in mitigating the fundamental frictions that make such systems essential. I also discuss the interactions between payments, monetary policy and financial stability. In the third section, I review ongoing changes that characterize payment systems nowadays. The last section concludes.

Correction 14 June 2019: Section 3.2 has been revised.

* I would like to thank Anders Vredin, David Vestin and Roger Broman for very useful comments and suggestions. The opinions expressed in this article are my sole responsibility and should not be interpreted as reflecting the views of Sveriges Riksbank.

1 The definition is from Roberds (2008).

2 The vast majority of payments in Sweden involve the transfer of bank deposits by various means.

3 Söderberg (2018) discusses in detail why the Riksbank was given a monopoly role to issue cash in Sweden.

1.1 Payment systems: historical evolution

Payments entail the transfer of value from one agent to another. When two agents exchange goods or services directly, such transfers are achieved by barter. There are, however, considerable frictions involved with barter transfers. Indeed, the Mengerian theory on the origins of money argues that the value of money is derived from the way it eases the friction of a 'double coincidence of wants' that hampers barter trade (Jevons 1875, Menger 1892). Menger's approach builds upon the idea expressed earlier by Jevons that for trade to take place in a barter economy, a consumer needs someone who not only has the desired good but also wants that consumer's good in return. In practice, it rarely occurs that two agents each want each other's good, still less that they have the correct quantities of each good available to be able to agree on the terms of their trade, and then still less that these coincidences materialize at the exact time that both sides of the bargain desire these goods.[4]

But while money has now acted as a mean of payment for a very long time, the documentary evidence suggests that, for a long time, payments among agents in the economy were for the most part limited to simple bilateral relationships – one agent would produce a good, and a consumer would pay for it with some form of money, either commodity or fiat.[5] It required the emergence of banks to create the conditions for the economics of payments to develop in a more sophisticated way.

Modern banks developed from different starting points. For example, bankers of the medieval Middle East engaged not only in money exchange and the granting of loans, but also regularly employed various payment methods. In Europe, even if merchants and bankers would have observed these practices when trading in and with the Muslim territories, there is no direct evidence to suggest that the Middle-Eastern payment instruments were directly adopted (Ashtor, 1973). Rather, in places where a wide variety of coins of different origin was in circulation, such as in thirteenth-century Venice, moneychangers expanded their role of valuing specie to offering payment and other banking services based on the deposits held with them (Kohn 1999, Mueller 1997).[6] Elsewhere, such as in mid-seventeenth century London, the origins of banking could be found among goldsmiths, who developed a similar banking business based on their specialist service of providing safekeeping facilities.

Whether western banks originated as moneychangers or as goldsmiths, merchants could deposit their coins with them in return for a receipt. Transactions could then be conducted either across the moneychangers/goldsmiths' books, or by transfer of the receipts they had issued. In some systems, such as those of continental Europe, account-based payment methods with transfers across accounts of banks tended to predominate. Elsewhere, such as in England (at least until the nineteenth century) or the US, issued notes were prevalent.[7]

By the start of the fourteenth century, Venetian records appear to show that account holders at the same bank could make payments to each other by book-entry transfer. There is no conclusive evidence, however, that these banks routinely accepted claims on each other. By the mid-fourteenth century, in the wake of a number of local bank failures, there were calls in Venice for a public bank to be set up, with the capacity to enable payments to be made without the credit risk that is inherent in commercial bank money. Its development

4 The Mengerian theory is complemented by the so-called Cartelist theory of the origins of money, which is based on the idea that money derives its value from the power and credibility of the issuing authority.

5 Cattle are the first and oldest form of money (9000–6000 B.C.). The first use of cowries, the shells of a mollusc that was widely available in the shallow waters of the Pacific and Indian Oceans, was in China (1200 B.C.). Bronze and copper cowrie imitations were manufactured by China and could be considered some of the earliest forms of metal coins (1000 B.C). For a history of money, see for example Davies (2016).

6 Specie is metallic money in all its forms (gold or silver traditionally).

7 Different payment approaches also reflected the different costs through time of addressing the vulnerabilities that were peculiar to each. With an account-based system, one must authenticate the account holder and keep records of the account holder's creditworthiness. Such a system is vulnerable to identity theft, and costly in record keeping. By contrast, with a store-of-value system (such as one based on issued bank-notes), one must verify the integrity of the store of value that is circulating. Over time, particularly with the technological advances of the twentieth century, the costs of account-based systems have decreased relative to the costs of store-of-value systems (see Kahn and Roberds, 2009).

was more than two centuries in the making and came to fruition only when the public Banco di Rialto was set up in 1587. But elsewhere in the Mediterranean trading area, municipal (so-called *Taula*) banks were set up as early as the start of the fifteenth century – including in Barcelona (in 1401), Genoa (1407) and Valencia (1408). The Taula enabled banks to hold deposits as reserves and to use these to clear interbank payments. However, even when the Banco di Rialto was imitated in other significant European trading cities – such as Amsterdam (1609), Hamburg (1619) and Nurnberg (1621) – the activity of these 'proto-central banks' was limited to local payments.

As for note-based systems, by the 1660s the London goldsmiths were carrying out a banking business in issuing notes against specie deposits, and creating money by issuing further notes to borrowers. The claims that banks accepted on each other were then redeemed on a bilateral basis every few days, with the (net) difference settled in specie. The frequency of settlement was determined by the creditworthiness of the issuer: the more reputable a banker, the longer other bankers were willing to hold his notes (Quinn, 1997).

As economic activity grew, following agricultural and industrial advances in the eighteenth and nineteenth centuries, increasing payments needed to be made over greater distances, and so both volumes and values of interbank obligations increased. In response, the banks' clearing and settlement arrangements became more formalized.[8] For instance, from 1775 onwards, the Bankers' Clearing House in London was settling daily. It was then a short step from settling obligations bilaterally, to doing so multilaterally and by 1841 onwards the Bankers' Clearing House in London started settling on a multilateral basis. The innovation of multilateral settlement further reduced the quantity of the settlement asset needed by participants to meet their net obligations.

A further cost-saving development, above all in note-based systems, was for the clearing-house blue-print to be adopted outside the capital cities. For example, in Canada, ten regional clearinghouses were similarly established between 1887 and 1902, with daily settlement at four main centres. Although such regional arrangements fragmented the pools of liquidity that the banks needed in order to settle their obligations (at least until regional net obligations were forwarded on to the centre), they saved significantly on settlement asset transportation costs.

These transportation costs were particularly acute when the settlement asset was specie or (gold) bullion. Not only was this cumbersome and costly to transport and exchange, but the process of transportation was vulnerable to a variety of operational risks – notably theft. To address the transportation costs, banks innovated by using assets that were convertible into specie and that all banks were willing to accept. By the 1770s, for instance, London's bankers had switched from settling in specie to settling in Bank of England notes. Notes issued by the Bank of England – as opposed to those issued by other London banks – were chosen presumably because of certain advantages uniquely enjoyed by the Bank, notably being the banker to the government.

Whether it was specie/gold or some paper that was (partially/wholly) convertible, a vulnerability to theft remained, however, as long as the settlement asset needed to be physically transported, either bilaterally between banks or to a (central) clearing house. By settling interbank obligations over accounts at a bank, this vulnerability could be eliminated entirely. In the United States, the Second Bank of the United States used its accounts to play an active role in providing inter-regional payment services for the two decades (1816–1836) that its charter was granted. And during the 1820s, Boston banks appointed a single agent for clearing and settling notes in Boston – the Suffolk Bank – which cleared the notes of several New England banks on a multilateral net basis and settled their positions in deposits that they held with it. The Suffolk Bank clearing and settlement arrangements were

8 Clearing is the process of reconciling and confirming payment orders prior to settlement, that is prior to the conclusion of a transaction.

superseded when, in the 1850s, the mutually owned Bank for Mutual Redemption replaced it and continued to perform a similar clearing and settlement function. Meanwhile, on the other side of the Atlantic, deposits at the Bank of England (as opposed to Bank of England notes) were used to effect settlement of the Bankers' Clearing House obligations from 1854 onwards.

This did not imply that one single institution should necessarily become the hub of a country's payment system. In Canada, for instance, regional clearing arrangements were formalized with the set-up (in statute) of a mutual Bankers' Association at the start of the twentieth century. Nevertheless, as technology constraints lessened (and specifically in the case of the Canadian system, as telegraphic communication between different regions became feasible), strong economies of scale emerged pushing for centralization. By 1927, settlement of Canadian banks' obligations was centralized at the Royal Trust Corporation.

Even where one institution did become the centre of a country's payment system, it was not necessarily the case that it would become the central bank. Continuing the Canadian case study, the Royal Trust Corporation was completely unrelated to the Bank of Canada that took over its settlement role when it was set up in 1935. Similarly, in post-Civil War US, a system of mutual clearing houses was set up. In 1913, the Federal Reserve System was established and started providing a unified nationwide interbank settlement system based on telegraphic wire transfers.

In Sweden, the Riksbank's involvement in clearing and settlement seems to have been limited between the mid-1850s and 1901.[9] Instead, the clearing function of the Swedish banking system was performed by two commercial banks. The Stockholms Enskilda Bank was started in 1856 and it immediately began to act as a clearing bank for other note-issuing banks. However, the Skandinaviska Kreditaktiebolaget, a non-issuing bank, largely took over the clearing function in the 1860s, mainly because it offered better terms. In 1897, a new bank law was promulgated that prescribed the monopolization of notes by the Riksbank. The transfer of the Enskilda banknotes to the Riksbank occurred between January 1901 and January 1904 and clearing was taken over by the Riksbank.

1.2 The advent of central banks' currency monopolies

Granting central banks the sole right to issue cash, known as the banknote monopoly, was a political decision taken in most of the Western world in the late nineteenth and early twentieth centuries. As discussed by Eichengreen (2019), monopoly over seigniorage is a source of political power and a valuable help when sovereignty is threatened. The various countries' reasons for introducing such monopolies differed, however, which makes it difficult to address the topic in a comprehensive manner. Thus, I will follow Söderberg (2018) and discuss the cases of the United Kingdom and the United States, the two largest Western economies at the time, as well as the one of Sweden.

The Bank of England was created in 1694 as a privately-owned bank that conducted lending operations with both the state and the general public in London. It also accepted deposits and issued paper banknotes. The second half of the 1810s in the United Kingdom was characterized by financial instability and inflation following the Napoleonic Wars. This led to a national discussion, which continued throughout the 1840s, on how to achieve a stable monetary and financial system. One conclusion of this debate was that inflationary pressures were due to the excessive issuance of banknotes by smaller banks. That is, the Bank of England could not control the quantity of banknotes in circulation and thus was not able to manage the total supply of money. As a consequence, the Bank Charter Act of 1844 heavily restricted the smaller banks' right to issue banknotes so that the Bank of England held the sole legal right to determine the number of banknotes in circulation.

9 Ögren (2006) argues that it is likely that the Riksbank did engage in some clearing activities.

In the United States, on the other hand, after 1836 and the dissolution of the Second Bank of the United States, a bank could be established without the permission of the states, provided that certain fixed capital requirements were met. Specifically, the banknotes issued by banks had to be redeemable against silver and gold and, in addition, banks had to allocate collateral in the form of federal or state bonds. By the mid-nineteenth century, there were over 1,500 private banks issuing banknotes in the United States (Gorton, 2012). These banknotes did not just circulate regionally, but also nationwide, which implied each bank had a large proportion of other banks' banknotes on its balance sheet (Rolnick et al., 1998). The underlying problem was that banknotes issued by different banks were not worth the same. Why? Banks had different risk profiles and several states suspended payments of their debts, thus undermining banks' collateral. The result was a highly impractical system in which traders had to determine how much different banknotes were worth in relative terms.[10] These difficulties, combined with the need to fund the American Civil War, which started in 1861, led to the National Bank Act of 1863 and the implementation of a system with federal, state-backed banknotes. The state banks' right to freely issue banknotes was eliminated and many of them were forced to close. National banks were created instead. These were privately owned banks, which could issue banknotes that were worth the same in all states and were backed by federal government bonds.

In the decades after the introduction of the national banknotes, no fewer than seven financial crises occurred in which bank runs were a central element (Gorton, 2012). The rationale for creating a central bank in the United States, with banknote monopoly, was therefore mainly provided by the need to create a lender of last resort (Wood, 2005). When the Federal Reserve was created in 1913, the decision was also taken to dismantle the national bank system and replace it with government banknotes issued by the Federal Reserve (Weyforth, 1925).

In Sweden, the Riksbank was the only bank for a long time. Banknote-issuing private banks were therefore allowed by the Swedish Riksdag in 1824 in order to promote the development of a banking system in Sweden. These banks accepted deposits from the general public, albeit on a very limited scale. Reasons for this included the limit placed on interest rates by older legislation on usury. Consequently, from the start, issuing banknotes was the private banks' main source of funding (Lilja, 2010). The system was, however, not entirely private. On the contrary, banknotes from the private banks could be redeemed for Riksbank banknotes, namely government banknotes, which could, in turn, be redeemed for precious metals. The relationship between the Riksbank and the private banks thus had strong similarities with an early central bank system. In addition, the banking system was under very strict government regulation. Permission to open a banknote-issuing bank was given by the central government and very restrictively. The legislation was also clearly formulated to limit what we today call moral hazard – banks should be organized like partnerships with unlimited economic responsibility and should not expect any government support in difficult periods (Jonung, 2007).

The profits from the issuance of banknotes, known as seigniorage, played an unusually important role in the debate on banknote monopoly that started in the 1840s in Sweden. The central government stood for a large part of the private banknotes' credibility, thus implying private banks received an indirect government subsidy. Consequently, arguments were made in Riksdag motions from the 1860s that the profit from the issuance of banknotes should belong to the government. Resistance to this argument was mainly justified by the argument that a monopoly would threaten the existence of Swedish banks (Brisman, 1931). The final decision was taken in 1897 and the Riksbank was granted monopoly power in the issuance of banknotes.

10 For example, a ten-dollar banknote issued in one state could be worth USD 9.90 in another state and USD 9.40 in a third (Gorton, 2012).

2 Brief review of the literature

While policy-makers care about the efficiency and stability of payment systems, guidance from economic theory has, until recently, been limited. This is changing – old models abstracted from the mechanism through which payments took place, whereas new models develop internally consistent, general-equilibrium models to analyse the roles of alternative payment instruments and institutions in facilitating trades. These are theories of rational, strategic agents, which explicitly model the underlying transactions of goods or financial assets that generate the use of the payment system.

The existing literature focuses on four key issues. First, it aims at identifying the fundamental frictions that underline the use of payments and settlement arrangements. Second, it investigates who should provide these systems, and what the government's role should be in mitigating the fundamental frictions. Third, it studies how payment systems interact with financial intermediation and macroeconomic policy. Last, it studies how these systems evolve with the ongoing improvements in information technology.[11]

The recent literature in monetary theory argues that *limited enforcement* and *limited information* are the two key micro-economic frictions that explain why some observed payment arrangements are essential to an economy. Limited enforcement refers to a situation where some agents can default on their obligations at little or no cost. Limited information refers to a situation where some agents have no or limited knowledge about other agents' current and/or past actions. In this environment, there may be a role for a central bank (and for regulation) with respect to mitigating some of these frictions through commitment and enforcement technologies. Several studies, however, find that it is not necessarily only a public agent who may fulfil these requirements for the welfare-enhancing provision of a payment system.

Indeed, theory generally suggests that central banks may have a comparative advantage in two main payment system functions. The first is the management of the accounts that participants own and use to settle transactions. Central banks are suited to this role because of their trustworthiness and public policy mandate. The second is the supply of very short-term credit (e.g., intraday credit) to intermediaries to facilitate settlement, or to facilitate the resolution of settlement disruptions. The provision of cheap central bank credit, however, may distort private sector choices by inducing participants to take excessive risks and overuse central bank credit, leading to moral-hazard problems. To deal with this issue, central banks are increasingly requiring collateral for such credit and liquidity and capital constraints can also be imposed.

A central bank may have several other advantages with respect to economies of scope, the commitment not to overissue the asset, to reabsorb the liquidity introduced by intraday needs, and to intervene in times of distress in view of its responsibility for price and financial stability. There are, of course, also examples where central banks have misused their privileged positions.

2.1 Optimal provision of payment systems

Green (2005) and Millard and Saporta (2005) point out several features that can make a central bank the best provider of payment systems, such as its neutrality with respect to financial institutions and its creditworthiness. They also argue that the provision of settlement accounts to banks is a natural extension of a central bank's traditional role as the government's banker. Their arguments are further strengthened if there exist economies of scope in the provision of the settlement asset and other public policy objectives, such as a concern about systemic risk that might not be internalized by private operators.

11 Chiu and Lai (2007) provide an extensive literature review on this topic.

Kahn and Roberds (2002) find that central-bank operated settlement arrangements may offer both potential advantages and disadvantages as compared to private ones. Although confidence in the liabilities of the central bank can sustain trade during crises, that same confidence can undermine the incentives of payment-system participants for mutual monitoring. This is of concern if one believes the public sector is less efficient at monitoring or is less inclined to act based on the information received. Thus, this disadvantage must be weighed when considering the merits of public versus private payment systems.

Mills (2004) studies whether outside money is needed for settlement. He finds this is not to be the case if a strong private enforcement authority exists. He conducts his analysis using a Freeman (1996a) model, who was the first to formulate a framework in which (i) debts are repaid with outside money and (ii) there can arise liquidity problems which create a role for a central bank discount window. Mills argues that outside money is not necessary for settlement in Freeman's environment – if agents have a technology to issue IOUs, and there is an enforcement authority to force debtors to redeem their own IOUs, then no outside money is needed for settlement. In the absence of this strong enforcement authority, however, outside money will be needed.

Green (1999) uses a Freeman (1996b) model to study whether efficiency might require a central bank to participate in the payment system. He finds this depends on the degree to which a central bank can promise reliably and credibly to reabsorb money that it issues to facilitate payments and on whether the commercial law framework governing the operation of a private-sector payment intermediary is enough to warrant agents' use of debt issued by the intermediary as a money-like medium of exchange. The credibility of a central bank's promise about reabsorption evidently depends, in turn, on its governance structure. Moreover, Green comments that it is likely that the institutions of central bank governance necessary for credible participation in the payment system are essentially identical to those necessary for the effective implementation of monetary policy in a narrow sense. Thus, to whatever extent there is a need for a central bank to participate directly in the payment system, this need reinforces the considerations in favour of chartering a politically independent central bank. Moreover, the need for political independence suggests that the central bank would typically be a more appropriate public-sector participant in the payment system than would the treasury or another agency under the immediate control of the government.

2.2 Central banks' currency monopoly

Currency monopoly is a controversial topic. Although significant numbers of economists opposed this development during its early stages (see Smith, 1990), others either favoured it or were indifferent. As monopoly became the norm, the opposition ceased – or did so until the mid-1970s, when Friedrich Hayek succeeded in reopening the debate, if only on a very small scale (see Hayek, 1976). Hayek advocated a system in which private entities would issue their own forms of money. Private monies would then compete among themselves to provide a stable means of exchange.[12]

As discussed by Selgin (2008), Milton Friedman's views on the issue offer a particularly interesting case study. Despite having been a firm supporter of free markets, he at first shared the common view concerning the necessity of official currency monopolies. In *Program for Monetary Stability* (Friedman, 1959), he asked the question 'whether monetary and banking arrangements could be left to the market, subject only to the general rules applying to all other economic activity.' 'Something like a moderately stable monetary framework' he wrote 'seems an essential prerequisite for the effective operation of a private market economy. It is dubious that the market can by itself provide such a framework.

12 Inside money (bank deposits) should not be confused with private money. A system of private money is one in which financial institutions create currencies that compete for acceptance. Cryptocurrencies can be viewed as examples of private monies.

Hence, the function of providing one is an essential governmental function on a par with the provision of a stable legal framework.' He revised his original opinions in light of the renewed interest in the question Hayek's work helped to stimulate. His opposition to banknote currency monopolies remained lukewarm, however, although he ultimately concluded there was, after all, 'no reason currently to prohibit banks or other groups from issuing hand-to-hand currency' (see Friedman and Schwartz, 1986).

The interest in privately-issued monies has been revived by the emergence of cryptocurrencies such as Bitcoin and Ethereum, but the monetary literature on this topic is still in its infancy. Among the few that have addressed this topic, Fernández-Villaverde and Sanches (2016) build an interesting model of competition among privately issued fiat currencies from which they derive three main insights. First, price stability can be consistent with competing private monies. Second, private monies are also subject to self-fulfilling inflationary episodes, even when they are issued by profit-maximizing, long-lived entrepreneurs who care about the future value of their monies. Third, a purely private monetary system does not provide the socially optimum quantity of money. That is, the market fails to provide the right amount of money in ways that it does not fail to provide the right amount of other goods and services.

2.3 Payment systems, financial stability and monetary policy

The efficiency and effective functioning of financial markets are affected by payment systems. For example, the instruments available for making payments, the clearing and settlement facilities to which financial market participants have access and whether there is a large-value transfer system (LVTS), among other things, greatly influence speed, financial risks, reliability and cost of transacting when financial market participants make payments.[13] Payment systems also contribute to integrating financial systems, both domestically and internationally.

On the downside, a payment system is one transmission mechanism through which unsound financial players can jeopardize the stability of the whole financial system, with potentially adverse effects on the real sector as well. As lenders of last resort, and in trying to ensure the stability of the financial system, monetary authorities may be forced to rescue individual banks and segments of the capital market to counter systemic risks to the financial system. The more fully integrated the financial markets and hence normally the more developed the payment system, the greater are the systemic risks that arise, underscoring the need for greater coordination of cross-border prudential measures to contain spillover effects.

Some of these threats may derive from the design and operation of payment systems themselves. Millard and Saporta (2007) identify two principal sources of systemic risk arising from payment and settlement activity: single point of failure risk; and risk arising from strategic interaction between payment system participants. In other words, faced with a prolonged disruption (or frequent disruptions) to the operation of a single provider of payment and settlement services in a particular market, users will be unable to re-route volume readily to an alternative provider. Trades may then remain unsettled for a period, either implying direct losses or creating unintended credit or market exposures.

In terms of monetary policy, central banks need to establish appropriate arrangements for liquidity provision to the banking system in order to effectively exercise control over the quantity and price at which its liabilities are made available. Central banks must therefore pay attention to the mechanisms by which they carry out such operations, ensuring the safety, resilience and efficiency of the payment and settlement systems used to mobilize

13 LVTSs are systems for electronic wire transfers of large sums of money.

collateral assets and distribute funds.[14] By extension, they must take an interest in the payment systems employed by the banking system to carry out its own credit intermediation and thereby transmit monetary policy more widely throughout the economy.

In recent years, there has been a marked shift away from deferred net settlement to real-time gross settlement (RTGS) in large-value payment systems.[15] This implies heightened intraday liquidity demands on payment system participants, which are typically met via the provision of credit by the central bank. Intraday credit is often extended against collateral, but typically at a very low nominal interest rate. Central banks' generous policy with respect to intraday liquidity provision has motivated a series of theoretical studies as to whether it is appropriate for a central bank to adopt different approaches to intraday versus overnight monetary policy.[16]

Millard et al. (2007) argue that the difference reflects a tension between different aspects of the central bank's monetary stability objective: on the one hand, a central bank sets overnight rates to meet its price stability objective; on the other, it may be prepared to inject intraday liquidity into an RTGS payment system at a very low cost to ensure that banks do not have an incentive to delay payments and risk settlement failure (as in Furfine and Stehm, 1998). Williamson (2009) argues that the distinction between intraday and overnight policy action has become blurred.

Changes in payment systems can occur because of reforms or innovation.[17] In that case, there are implications for the monetary-policy decision making process that go beyond the need to consider the impact on the demand for base money. According to Johnson et al. (1998), four different areas of decision making could be affected in this case. First are the *monetary policy target and instrument settings* – for example, the aggregate volume of reserves the central bank should supply for consistency between payment-related demand for reserves and the central bank's desired monetary policy stance; the pricing or the quantity limits in standing central bank credit facilities, and the appropriate relationship between very short-term interbank interest rates (which the central bank directly affects) and other interest rates and financial variables (over which the central bank has less direct influence). Second are the *choice and interpretation of appropriate target or indicator variables* for monetary policy, at least during some transitional period – for example, the relative weights (or reliability as indicators) attached to price and quantity variables (interest rates versus reserve money) while demand for the key operational quantity variable (reserve balances) is shifting. There may be effects on quantity variables at the level of the banking system, as well as at the level of the central bank's balance sheet, to the extent that, for instance, the velocity of transaction balances is altered. Third is the appropriate *design of monetary policy instruments* – for example, the design of reserve requirements or central bank standing credit facilities, or the nature and timing of central bank market operations might need to be adjusted considering payment system reforms or endogenous changes. Fourth, of course, is the *monetary policy transmission mechanism* itself – for example, the efficiency with which central bank changes in the supply of reserves affect interest rates in different markets and hence other economic and financial variables of ultimate interest.

14 Difficulties for monetary policy can arise when inefficiencies and changes in the payment system cause unpredictable shifts in the demand for base money. Examples of such inefficiencies include long delays in processing and settling payments, payment gridlock, frequent breakdowns in payment facilities, large-scale fraud, stoppage in operations and lack of clarity over important legal issues affecting payments, such as bankruptcy laws, legality of documents in contracting and enforceability of contracts and agreements.

15 With deferred settlement systems, final settlement of transfer instructions occurs on a net basis at one or more discrete, pre-specified times during the processing day. With RTGS systems, final settlement of interbank funds transfers occurs on a continuous, transaction-by-transaction basis throughout the processing day.

16 See the proceedings from the 2007 conference on 'Payments and Monetary and Financial Stability' organized by the European Central Bank and the Bank of England.

17 Recent examples of payment innovations are cryptocurrencies such as Bitcoin and the use of blockchain technology for inter/intrabank payments (see Prasad, 2018).

3 Ongoing changes in payment systems

Despite their central role in the economy, payment systems have long been considered only the plumbing of the economy, meaning something essential but boring. Recently, however, they have received more attention, due to the significant technological changes in the industry and associated policy concerns, and possibly simply because of the magnitude of payment activity. In this section, I will touch upon some of the issues of key importance that central banks, and in particular the Riksbank, face nowadays when it comes to payment systems.

3.1 The role of cash

In most Western nations, legal tender status has typically been extended to coins and banknotes, meaning the governments of those countries consider them to be the official money in use. Nowadays, however, money is not just banknotes, but it takes many different forms: debit cards, cheques, and contactless payments using mobile devices, among others. Moreover, in retail situations, it is not uncommon to encounter shopkeepers refusing banknotes. For example, many US and EU retailers refuse to accept $100 and €500 bills respectively and many Swedish shops have dispensed with cash altogether. This can happen because of how the interaction between the legal tender provision and the principle of freedom of contract is governed in those countries.

In Sweden, for example, banknotes and coins issued by the Riksbank are legal tender (see Chapter 5, Article 1 of the Sveriges Riksbank Act). However, a principle of freedom of contract applies in the country which implies contract laws have a higher precedence than banking and payment laws here. In other words, the parties are free to agree on another method of payment than cash, and the provision of legal tender simply establishes a right to pay in cash unless otherwise agreed between the parties. This implies that if a store puts up a sign that it does not accept cash, then a customer has entered a contract or an agreement with that store that it does not accept cash.

Quite similarly, in the United States, section 31 U.S.C. 5103 states: 'United States coins and currency [including Federal Reserve notes and circulating notes of Federal Reserve banks and national banks] are legal tender for all debts, public charges, taxes, and dues.' This means that all United States currency constitutes a valid and legal offer of payment for debts to creditors. There is, however, no Federal rule that a private business, a person, or an organization must accept currency or coins as payment for goods or services. Private businesses are free to develop their own policies on whether to accept cash unless there is a state law stating otherwise.

In the Euro area, the status of euro banknotes and coins as legal tender follows from Article 128 of the 2010 Treaty on the Functioning of the European Union and Council Regulation no. 974/98. Also in 2010, the European Commission issued a recommendation (2010/191/EU of 22 March 2010) on the scope and effects of legal tender of euro banknotes and coins. The recommendation includes a number of guiding principles, including: '1. The acceptance of euro banknotes and coins as means of payments in retail transactions should be the rule. A refusal thereof should be possible only if grounded on reasons related to the "good faith principle" (for example, the retailer has no change available). 2. Large banknotes should be accepted as means of payment in retail transactions unless refusal thereof is grounded on reasons related to the "good faith principle" (for example, the face value of the banknote tendered is disproportionate compared to the amount owed to the creditor of the payment). 3. No surcharges should be imposed on payments with euro banknotes and coins.' Such recommendations are non-binding however and four member states (Germany, Finland, the Netherlands and Ireland) did not endorse them. Indeed, according to these four member states, the principle of contractual freedom can limit legal tender provisions.

In Denmark, instead, the principle of freedom of contract does not supersede the country's so-called *cash rule*, originally passed in 1984. Under such rule, a payee who accepts payment instruments, such as payment cards, must accept cash as payment in connection with staffed sales (section 56 of the Payment Services Act).[18] This means that cash cannot be rejected whenever the retailer is staffed. On the other hand, the payee does not have any obligation to accept cash in connection with online purchases and in unstaffed self-service environments such as unstaffed gas stations. The cash rule applies to both private sector firms and public institutions.

This is an important issue, especially for central banks of countries where the use of cash is declining. In countries like Sweden, for example, where such decline is particularly striking (see Sveriges Riksbank, 2018a), one can easily envisage a near future in which cash will de facto not be accepted. This could be true even if cash were still legal tender, the reason being that cash acceptance is intimately connected with the expectation that it will be valuable in the future. Its value therefore declines if it is accepted less frequently, as shown by Bigoni et al. (2018). Central banks should thus investigate how to react to that possibility and whether cash should be viewed as the ultimate mean of settlement, meaning that it could function even in the event of blackouts or wars. As such, should its acceptance be guaranteed, much like it happens in Denmark? And if that were to be the case, would competition in the payments industry be guaranteed?

Partly as a response to the decline role of cash, several central banks have publicly announced internal efforts to study central bank digital currencies (CBDCs) for either retail payments or wholesale payments, or both. The Riksbank is at the forefront of this debate (see Sveriges Riksbank, 2018b). If implemented, the impact of CBDCs would be significant: banks have traditionally played a central role in supporting payments, so that removing them from the centre of this system could reshape banking and, more broadly, the financial markets. Such potential important consequences are currently the subject of debate among economists (e.g. see Andolfatto 2018, Bech and Garratt 2017, Berentsen and Schar 2018, Bordo and Levin 2017, Cecchetti and Schoenholtz 2017, among others).

3.2 Fast-payment initiatives

An important distinction must be made between wholesale and retail payments. Wholesale payments are high-priority and typically large-value transfers that are made between financial institutions for their own accounts or on behalf of their customers and are usually settled via dedicated interbank settlement systems. Retail payments, instead, are lower-value transactions between individuals, businesses and governments in such forms as cash, cheques, credit transfers, and debit and credit card transactions (see Bech et al., 2017 for an interesting discussion).

The speed of settlement finality has traditionally been very different between the two. Traditional retail payment systems were typically slower, and for some systems, payments were revocable within a certain period. Hence, time-sensitive payments (even lower-value ones) were directed via the interbank payment system because of its ability to credit and debit accounts with real-time finality. Nowadays, however, the speed of retail payments is immediate in some countries thanks to improvements in information and communication technologies (e.g. smartphones and the internet).

In Sweden, **Bankgirot** is a bank-owned clearing organisation and is the central actor in the mediation of retail payments in the country. Several different types of payments and transfers are made through Bankgirot, such as credit transfers, direct debits, suppliers' payments, salary payments, account-to-account transfers, and the clearing of card payments and ATM withdrawals. The mobile payment system **Swish**, which was introduced in 2012,

18 See Danish Payments Council (2018), 'Report on the Role of Cash in Society', for details.

is the first service to use Bankgirot's settlement system for real-time payments. Thus, Swish allows private individuals access to a real-time payment service. Indeed, the typical time between payment initiation and availability of final funds to the payee for a successful payment transaction is one to two seconds.

As for wholesale payments, in recent years there has been a marked shift away from deferred net settlement to real-time gross settlement (RTGS), as discussed in Section 2.3. Many RTGS systems are currently in the process of being renewed or were recently renewed. Reasons for this include technological innovation, the end of technological lifecycles, the emergence of new players and regulatory standards, and the growing prominence of fast-payment initiatives. Overall, key RTGS design parameters are evolving to accommodate more access by non-banks, greater interoperability between payment systems and longer operating hours.

In Sweden, **RIX** is the payment system for large-value payments and the hub of the Swedish financial infrastructure, as RIX deals with payments to and from banks' accounts. Transfers between the accounts held by participants in RIX are made electronically and according to the RTGS principles. The system is owned and operated by Sveriges Riksbank, which can provide credit to RIX participants during the day, what is known as 'intraday credit'. However, credit is only granted against adequate collateral.

In Europe, the ECB launched a new RTGS system in 2018: TARGET Instant Payment Settlement (TIPS). As settlement in **TIPS** takes place in central bank money, participation in TIPS depends on being eligible to access such funds. TIPS currently only settles payment transfers in euro. However, in case of demand other currencies could be supported as well.

TARGET2-Securites (T2S) is also an RTGS platform on which securities and payments can be transferred simultaneously between investors across Europe, using harmonized rules and practices. T2S thus lays the foundations for a single market for securities settlement and contributes to achieving greater integration of Europe's financial markets. Banks pay for securities on the platform using the account they have with their central bank, so the money used to settle transactions is central bank money. As a result, transaction risk is greatly reduced.

These developments bring to the forefront two fundamental issues: the optimal role of central bank money in the payments chain and the access policy for RGTS systems. First, while most payments are eventually settled in central bank money, how much of these flows should be settled on a gross basis using central bank money as opposed to commercial bank money is open to debate. Second, direct access to RTGS systems has typically been limited to entities such as banks. Over the years, however, demand for access has increased from new entrants into the payments business and this trend is expected to continue (see Bech et al., 2017). Potential changes to access criteria bring up competition and financial stability considerations for central banks. On the one hand, limiting access to RTGS systems to banks could unduly favour these entities over others and potentially limit innovation in the market for payments. On the other hand, if access is extended to wider entities such as non-banks, central banks and regulatory authorities will have to consider how to guarantee equal opportunities for all players in aspects such as regulatory and supervisory frameworks and reserve requirements. In addition, wider access could have an impact on the overall resilience of the RTGS system. On the one hand, new entrants could be detrimental to system resilience if they are not as strictly supervised as banks. On the other hand, wider access could also have a positive impact on the overall stability of payment systems if the central bank strengthens the monitoring of new players as a necessary condition of their access to the RTGS system.

4 Concluding remarks

Payment systems are currently undergoing important changes, mostly because of technological innovations. Such changes include a declining role for cash and a growing prominence of fast-payment solutions, both of which create new challenges for central banks. For example, should cash be viewed as the ultimate mean of settlement and should its acceptance be guaranteed by law? If so, how do we guarantee competition in the payment system sector? And who should have access to real time gross settlement (RTGS) platforms given the new players in the field? In order to understand the best role of central banks in this new environment, we must first understand how money and payment systems in general came to acquire the role they have today and what the literature in this field can teach us.

References

Andolfatto, David (2018), 'Assessing the Impact of Central Bank Digital Currency on Private Banks', St. Louis Fed Working Paper no. 2018-026C.

Ashtor, E. (1973), 'Banking Instruments between the Muslim East and the Christian West', *Journal of European Economic History* vol. 1, pp. 553–573.

Bech, Morten L., Yuuki Shimizu and Paul Wong (2017), 'The Quest for Speed in Payments', *The BIS Quarterly Review*, March.

Bech, Morten L. and Rodney Garratt (2017), 'Central Bank Cryptocurrencies', *The BIS Quarterly Review*, September.

Berentsen, Aleksander and Fabian Schar (2018), 'The Case for Central Bank Electronic Money and the Non-case for Central Bank Cryptocurrencies', *Federal Reserve Bank of St. Louis Review* vol. 100, no. 2, pp. 97–106.

Bigoni, Maria, Gabriele Camera and Marco Casari (2018), 'Partners or Strangers? Cooperation, Monetary Trade, and the Choice of Scale of Interaction', *American Economic Journal: Micro* (forthcoming).

Bordo, Michael D. and Andrew T. Levin (2017), 'Central Bank Digital Currency and the Future of Monetary Policy', NBER Working Paper no. 23711.

Brisman, Sven (1931), 'Den Stora Reformperioden 1860–1904 [The Great Reform Period 1860–1904]', In *Sveriges Riksbank 1668–1924: Bankens Tillkomst och Verksamhet* [Sveriges Riksbank 1668–1924: The Bank's Origin and Activities], vol. 4, P.A. Norstedt & Söner: Stockholm.

Cecchetti, Stephen G. and Kim L. Schoenholtz (2017), 'Fintech, Central Banking and Digital Currency', *Money and Banking Blog*, June 12, 2017.

Chiu, Jonathan and Alexandra Lai (2007), 'Modelling Payments Systems: A Review of the Literature', Staff Working Paper no. 2007-28, Bank of Canada.

Danish Payments Council (2018), 'Report on the Role of Cash in Society'.

Davies, Glyn (2016), 'A History of Money from Ancient Times to the Present Day', University of Wales Press: Cardiff, 4th edition.

Eichengreen, Barry (2019), 'From Commodity to Fiat and Now to Crypto: What Does History Tell Us', NBER Working Paper no. 25426.

Fernández-Villaverde, Jesús and Daniel Sanches (2016), 'Can Currency Competition Work?', NBER Working Paper no. 22157.

Freeman, Scott (1996a), 'The Payments System, Liquidity, and Rediscounting', *American Economic Review* vol. 86, no. 5, pp. 1126–1138.

Freeman, Scott (1996b), 'Clearinghouse Banks and Banknote Over-issue', *Journal of Monetary Economics*, vol. 38, pp. 101–115.

Friedman, Milton (1959), *A Program for Monetary Stability*, Fordham University Press: New York.

Friedman, Milton and Anna J. Schwartz (1986), 'Has Government Any Role in Money?', *Journal of Monetary Economics* vol. 17, pp. 37–62.

Furfine, Craig H. and Jeff Stehm (1998), 'Analysing Alternative Intraday Credit Policies in Real-time Gross Settlement Systems', *Journal of Money, Credit and Banking* vol. 30, no. 4, pp. 832–848.

Gorton, Gary B. (2012), *Misunderstanding Financial Crises*, Oxford University Press: Oxford.

Green, Edward J. (1999), 'Money and Debt in the Structure of Payments', *Federal Reserve Bank of Minneapolis Quarterly Review* (Spring), pp. 13–29.

Green, Edward J. (2005), 'What is the Role of a Central Bank in Payment Systems?', Pennsylvania State University Working Paper.

Hayek, Friedrich (1976), *Denationalisation of Money: An Analysis of the Theory and Practice of Concurrent Currencies*, The Institute of Economic Affairs.

Humphrey, David, Magnus Willesson, Göran Bergendahl and Ted Lindblom (2006), 'Benefits from a Changing Payment Technology in European Banking', *Journal of Banking and Finance* vol. 30, pp. 1631–1652.

Jevons, William S. (1875), *Money and the Mechanisms of Exchange*, D. Appleton & Company, New York.

Johnson, Omotunde, Jean-Marc Destresse, Nicholas Roberts, Mark Swinburne, Tonny Lybek and Richard K. Abrams (1998), 'Payment Systems, Monetary Policy and the Role of the Central Bank', IMF, Washington, DC.

Jonung, Lars (2007), 'Private Bank Notes in Sweden 1831–1902', Stockholm School of Economics manuscript.

Kahn, Charles M. and William Roberds (2002), 'Payments Settlement under Limited Enforcement: Private versus Public Systems', Federal Reserve Bank of Atlanta Working Paper no. 2002-33.

Kahn, Charles M. and William Roberds, (2009), 'Why Pay? An Introduction to Payment Economics', *Journal of Financial Intermediation* vol. 18, no. 1, pp. 1–23.

Kohn, Meir (1999), 'Early Deposit Banking', Dartmouth University Working Paper no. 99-03.

Lilja, Kristina (2010), 'The Deposit Market Revolution in Sweden', In *The Swedish Financial Revolution*, ed. by Ögren, Anders, Palgrave: Basingstoke.

Manning, Mark, Erlend Nier and Jochen Schanz (2009), *The Economics of Large-Value Payments and Settlement: Theory and Policy Issues for Central Banks*, Oxford University Press, New York, USA.

Menger, Carl (1892), 'On the Origins of Money', *Economic Journal* vol. 2, pp. 239–255.

Millard, Stephen and Victoria Saporta (2005), 'Central Banks and Payment Systems: Past, Present and Future', Bank of England manuscript.

Millard, Stephen, George Speight and Matthew Willison (2007), 'Why do Central Banks Observe a Distinction between Intraday and Overnight Rates?', Bank of England manuscript.

Mills, David (2004), 'Mechanism Design and the Role of Enforcement in Freeman's Model of Payments', *Review of Economic Dynamics* vol. 7, pp. 219–236.

Mueller, Reinhold C. (1997), *The Venetian Money Market – Banks, Panics, and the Public Debt 1200–1500*, Johns Hopkins University Press, Baltimore, MD, USA.

Norman, Ben, Rachel Shaw and George Speight (2011), 'The History of Interbank Settlement Arrangements: Exploring Central Banks' Role in the Payment System', Bank of England Working Paper no. 412.

Ögren, Anders (2006), 'Free or Central Banking? Liquidity and Financial Deepening in Sweden, 1834–1913', *Explorations in Economic History* vol. 43, no. 1, pp. 64–93.

Prasad, Eswar (2018), 'How will FinTech and Digital Currencies Transform Central Banking', Hutchins Center Working Paper, Brookings Institution.

Quinn, Stephen (1997), 'Goldsmith-banking: Mutual Acceptance and Interbank Clearing in Restoration London', *Explorations in Economic History* vol. 34, no. 4, pp. 411–432.

Roberds, William (2008), 'Payment Systems', In: Durlauf S. N., Blume L. E. (eds), *The New Palgrave Dictionary of Economics*, Palgrave Macmillan, London.

Rolnick, Arthur J., Bruce D. Smith and Warren E. Weber (1998), 'Lessons from a Laissez-faire Payments System: the Suffolk Banking System (1825–58)', *Federal Reserve Bank of St. Louis Review* (May/June).

Selgin, George (2008), 'Milton Friedman and the Case against Currency Monopoly', *Cato Journal* vol. 28, no. 2, pp. 287–301.

Smith, Vera C. ([1936] 1990), *The Rationale of Central Banking*, Liberty Press: Indianapolis.

Söderberg, Gabriel (2018), 'Why did the Riksbank get a Monopoly on Banknotes?', *Sveriges Riksbank Economic Review* no. 3, pp. 6–15.

Sveriges Riksbank (2018a), 'Payment Patterns in Sweden 2018'.

Sveriges Riksbank (2018b), 'Special Issue on the E-Krona', *Sveriges Riksbank Economic Review* no. 3.

Weyforth, W.O. (1925), 'The Retirement of National Bank Notes', *Journal of Political Economy* vol. 3, no. 5, pp. 531–549.

Williamson, Stephen D. (2009), 'Transactions, Credit, and Central Banking in a Model of Segmented Markets', *Review of Economic Dynamics* vol. 2, pp. 344–362.

Wood, John H. (2005), *A History of Central Banking in Great Britain and the United States*, Cambridge University Press: Cambridge.