Introduction

'We'd always thought that if you wanted to cripple the US economy, you'd take out the payment systems. Banks would be forced to fall back on inefficient physical transfers of money. Businesses would resort to barter and IOUs; the level of economic activity across the country could drop like a rock.'

Alan Greenspan, Chairman of the Board of Governors of the US Federal Reserve System, 1987–2006.

Today's complex web of market infrastructures – payment, clearing and settlement systems – is a response to frictions that arise when goods and financial securities are traded.

In general equilibrium models, developed famously by Arrow and Debreu (1954), such frictions are assumed away. Every agent can trade costlessly with everyone else, while the 'Walrasian auctioneer' ensures an allocation that matches individual needs. Each agent immediately finds a trading partner, there is no need for money, and no role for either central banks or payment systems to support the flow of money through the economy. Moreover, in the Arrow–Debreu world, agents can commit to contracts, ensuring that all obligations will be honoured. There is, therefore, no need for sophisticated mechanisms to enforce commitment or manage the risks associated with non-performance.

Sadly, in the real world, trading frictions *do* exist. There is then a role for money, a role for central banks, and a role for a payment and settlement infrastructure that both supports the flow of money and provides mechanisms for management of the risks associated with limited enforcement of contracts.

With a view to clarifying policies to address risks in payment systems, central banks have over the past few years devoted considerable resources to the study of the economics of payments. In parallel, this field has begun to establish itself as a new subject for scholarly research, drawing in academic students and researchers. To date, however, there has been little attempt to draw together the key insights gained from this growing body of research.

This volume seeks to offer just such a synthesis. It charts the frontier of our knowledge to date and puts it in the context of a comprehensive overview of the policy issues faced by central banks in this sphere. In particular, it explores: central banks' roles in payment systems; the risks on which central banks focus in their oversight activities; and the challenges central banks face as the payments and settlement landscape evolves.

The economics of payments is a multidisciplinary field, taking in branches of economics such as monetary theory, search theory, game theory and industrial organization. It also draws on techniques from network theory and makes extensive use of simulation studies to model complex interactions between payment-system members. For each of the topics covered, this volume highlights some of the most influential works in the literature. The volume also draws heavily on empirical insights, in particular offering an historical context to central banks' involvement in payment systems.

We hope that this book will appeal to a wide audience, including fellow policymakers, practitioners – either users or providers of infrastructural services – and non-specialist academic economists and students. Not only do we hope it will offer a good reference text, but also that it will stimulate active debate and encourage new research in this important field.

1 Money, banking and payments: historical evolution and the role of the central bank

What is a payment? One starting point is to consider the etymology of the verb 'to pay': it is derived from the verb 'to pacify'. In this context, the idea of pacifying goes back to an age before modern systems of payments had been developed: the concept is closely connected, for instance, to the medieval practice of *wergild*, whereby a guilty party paid a fine to its

victims in order to prevent blood feuds. In today's usage, a payment still describes a transfer of value between agents.[1] By extension, any *organized* arrangement for transferring value between parties can be defined as a payment *system*. Thus, the Bank for International Settlements (CPSS, 2003c) defines a payment system as 'a set of instruments, banking procedures and, typically, interbank funds transfer systems that ensure the circulation of money.'

Such systems developed in the context of the broader evolution of money and banks and, eventually, central banks. In **Part I** of this volume, we explore this evolution, describing how central banks came to assume a role at the heart of the payments infrastructure, and illustrating why it is natural for this role to be combined with a responsibility for monetary and financial stability. We show how the role of central banks is intimately linked to the evolution of 'large value' or 'wholesale' payment systems that facilitate the flow of liquidity between banks.[2]

The early banks emerged from the business of moneychangers and goldsmiths. These early banks developed the first arrangements for making payments in-bank; that is, they enabled merchants to make payments to one another without having to hand over specie (e.g. gold and silver coins), on the basis either of deposits held with the same bank or of depositor receipts (effectively bank-notes). Over time, so as to accommodate transfers between customers of different banks, the banks started accepting claims on each other.

Once banks started building up claims on each other, they needed, at some point, to be able to extinguish or 'settle' these claims. Such 'final settlement' had to occur via the transfer of an asset that the creditor bank was prepared to accept: the ultimate settlement asset. At first, banks started settling interbank claims using specie. Later, they switched to settlement in assets convertible into specie: for example, by the 1770s, London bankers had switched from settling in specie to settling in Bank of England notes, deemed a superior form of settlement asset. By the mid-nineteenth century, banks had innovated further, and were able to settle claims across the Bank of England's books.

[1] And in this connection it is worth noting that a payment today is, like the *wergild* of Anglo-Saxon times, a transfer of value from one agent to another, i.e. a one-way transaction.

[2] Given our focus on wholesale systems and financial stability, we do not review the economics of retail payment systems, such as credit cards, where the literature on network effects and two-sided markets has been successfully applied to analyze competition. A survey of this literaure can be found in Rochet and Tirole (2004).

Elsewhere too, such as in the US or Canada during the nineteenth and early twentieth centuries, a single, central institution emerged at which banks could settle their claims. But this institution was not necessarily a bank – as the clearinghouse model of settlement demonstrated – nor did it necessarily issue the ultimate settlement asset. The connection between central clearing and settlement and central banking was therefore neither straightforward nor uniform in its evolution – at least not until the twentieth century, when several governments set-up central banks that, in providing the ultimate settlement asset, took over (or, in cases where it did not already exist, established) the central settlement function. During the course of the twentieth century, these central banks then also assumed the two key central-banking functions of ensuring monetary and financial stability.

With its liabilities used as the ultimate settlement asset, a central bank has an incentive to maintain their value by setting the terms on which they are made available to the banking system. There is therefore a clear synergy between a central bank's roles in providing the settlement asset and promoting monetary stability. Furthermore, to ensure that balances held in bank deposits – by far the largest component of the money supply in a modern monetary economy – can continue to function as a medium of exchange, the central bank will take a natural interest in the payment systems employed in the transfer of bank deposits. Indeed, a payment system that enables the transfer of commercial bank money and that utilizes central-bank money as its settlement asset is crucial for the implementation and transmission of monetary-policy decisions taken by central banks.

Similarly, there is a link between the provision of the ultimate settlement asset and a central bank's financial stability objective. The provider of the ultimate settlement asset can, in times of crisis, increase the supply of its liabilities to ensure that payments continue to be settled. Under certain circumstances, it might also effect emergency lending to prevent the failure of a solvent but illiquid institution.[3] However, in order to avoid moral hazard, there are benefits to reducing threats to the financial system *ex ante* if this can avoid the need to intervene *ex post*. Indeed, such threats may arise from weaknesses in the design or operation of the payment systems themselves.

[3] This capability is often described as a central bank's 'lender of last resort' function.

Today, central banks around the world still typically provide the ultimate settlement asset, at least for large-value and wholesale market payments, but often for major retail systems also. They sometimes also operate, or even own, key components of the payments and settlement infrastructure.

But over time the financial infrastructural landscape has become more complex, expanding well beyond the traditional domain of central banks. New payment systems and other clearing and settlement infrastructures have emerged, processing particular categories of payments and supporting post-trade processing in particular financial markets.

And monetary economies have become increasingly dependent upon the existence of this machinery, with values and volumes passing through core infrastructures rising rapidly over time. Central banks have, therefore, taken an active interest, typically via the assumption of an oversight role, in ensuring that the infrastructure more broadly is operationally effective, efficient and resilient, and that undue risks are not imposed upon system members and the financial system more generally.

That is, they have sought to ensure that the design and operation of the infrastructure itself does not pose a risk to financial, or systemic, stability. Systemic risk, in this context, may be defined as 'the risk that the failure of one participant in a transfer system, or in financial markets generally, to meet its required obligations will cause other participants or financial institutions to be unable to meet their obligations (including settlement obligations in a transfer system) when due.'[4]

2 Sources of systemic risk in payments and settlement

But what exactly could go wrong? This is considered in **Part II** of this volume. In payment systems, the key source of systemic risk is settlement risk: that is, the risk that settlement will 'not take place as expected' (CPSS, 2003c). Settlement risk has its roots in four broad categories of risk: credit risk; liquidity risk; operational risk; and business risk. These may be defined, as follows:

- *Credit risk*: the risk that a participant in the system defaults on its obligations within the payment system, imposing direct unanticipated losses on other members;

[4] CPSS (2003c).

- *Liquidity risk*: the risk that a participant/participants hold(s) insufficient liquidity in the settlement asset, disrupting the flow of liquidity in the system and leading to delay in or failure of its and other participants' settlements;
- *Operational risk*: the risk of losses arising from technical failure or other forced interruption to the operations of a payment system (or its core components), or those of its participants;
- *Business risk*: the risk of losses arising from either suspension (or termination) of a payment system's (or its core components') provision of services due to financial pressures.

Notice that some of these risks affect participants and others the system itself. Credit and liquidity risks, for instance, are a characteristic of the interconnections and strategic interactions between the banks involved and would exist whether or not they were members of a payment system. Where there is a system, however, it can be the channel through which these contagious effects are manifest. Whether the system transmits these effects depends on its design. Operational and business risk, on the other hand, can occur both at the level of the participant and at the level of the system.

2.1 *Sources and mitigation of credit risk in payment systems*

We first home in on credit risk, exploring how this risk might manifest itself and identifying the specific threats it poses to financial stability.

Credit risk in payment and settlement systems can be influenced significantly by payment system design and, in particular, the frequency of settlement. Historically, interbank payments were typically settled on a net basis. Under this mode of settlement, payment requests received by the banks are collected together over a period of time and the net amounts that the banks owe/are owed are calculated. At the end of the period, the banks settle these net amounts by paying the settlement asset into the system (if they owe) and receiving the settlement asset from the system (if they are owed).

But such systems can transmit individual bank failures more widely. This is because, if payments are credited to customer accounts before final settlement occurs, credit exposures between banks can build up and a failure of one participant in the system can lead to the failure of others.

In an early simulation study,[5] there was found to be significant scope for systemic spillover in the US CHIPS system, which, at that time, settled on a deferred net basis. Subsequent studies, building on this methodology and applying it to other systems, have identified less scope for spillover.

Nevertheless, over the past two decades, in part driven by public authorities, there has been a marked shift towards real-time gross settlement (RTGS) of (at least) wholesale, interbank payments. Under RTGS, each payment instruction submitted to the system is settled individually with finality in real time intraday, eliminating the interbank credit risk that arises from a delay in settlement. Similar considerations also prompted the adoption of delivery-versus-payment (DvP) in securities settlement and, more recently, payment-versus-payment (PvP) in the settlement of foreign-exchange transactions. The benefit of these risk-mitigating measures has been evidenced during the market turbulence of late 2007 and 2008, with settlement able to proceed smoothly across markets, notwithstanding the prevalence of counterparty credit concerns.

2.2 *Sources and mitigation of liquidity risk in payment systems*

RTGS systems do, however, carry risks of their own. In particular, such systems are 'liquidity hungry' relative to net systems. That is, participant banks require more liquidity to settle their payments in a timely manner. With RTGS now widely adopted, the focus of attention has therefore turned to liquidity risk.

Since holding liquidity may be costly, banks participating directly in an RTGS system have an incentive to delay non-time-critical payments in the expectation that incoming receipts will provide the liquidity for their outgoing payments. But if all banks are doing this, there is a risk of 'gridlock' in the system: the risk that all participants are relying on incoming liquidity, but no-one is in a position (or willing) to make the first payment. This strategic interaction has been the subject of considerable academic and policy interest in recent years.

In particular, it is important that appropriate system-design features are incorporated to manage the liquidity burden on member banks that arises in RTGS systems and to prevent either disruption to the flow of liquidity within the system or the diversion of flows to less robust vehicles/systems. To reduce the likelihood of liquidity shortages, central banks may decide to provide intraday liquidity on generous terms, for instance by charging a

[5] Humphrey (1986).

low (or zero) interest rate, and by accepting a broad range of securities as collateral. Indeed, with internationally active banks increasingly participating in multiple payment systems, there have been strong industry demands for increased central-bank acceptance of collateral denominated in foreign currency.

The design of the system might also be adapted to reduce the liquidity burden of RTGS, while not compromising on credit risk. In this regard, so-called *hybrid* systems have emerged in recent years, with in-built functionality to queue certain outgoing payments until offsetting incoming payments have arrived.

2.3 *Operational and business risk in payment systems*

While the literature has tended to focus primarily on credit and liquidity risk, it is clear that other sources of settlement risk – namely, operational and business risks – might also be important, particularly with concentrated provision of infrastructure, where 'single-point-of-failure' issues loom large.

The operational capacity of the central infrastructure and/or member settlement banks to process payments in the normal way could be compromised by a number of factors, both internal – e.g. technical failures sourced within the system itself or the way in which it is operated – and external – e.g. problems with a source external to the system such as general power failures, terrorist action or natural disasters.

Appropriate technical-level system design, accompanied by sufficiently robust operational procedures, should serve to minimize the likelihood of a failure with a source internal to the system itself. Effective contingency arrangements and workarounds – back-up facilities; contingency sites, etc. – might then reduce the impact of those incidents that do, nevertheless, occasionally occur.

But it is not only the ability of the central system to process payments that is crucial here: the effectiveness of liquidity recycling within the system also relies on uninterrupted operations at the level of the settlement banks. In the extreme, the inability of a settlement bank to send payments raises the possibility of a 'liquidity sink' developing in an RTGS system, as available liquidity becomes concentrated in the settlement account of the bank concerned. Intraday, this could cause liquidity shortages elsewhere in the system, which might in turn lead to significant

delays to the settlement of payments between the unaffected settlement banks.

The likelihood that a member-level disruption will undermine the effectiveness of liquidity recycling in the system as a whole depends crucially on the structure of the payments network and the behavioural responses of other banks in the system.

Business risk, on the other hand, is the risk that the payment system provider becomes insolvent. A variety of tools may be applied to address this risk in payment systems. Of these, perhaps the most obvious is the application of capital requirements to payment systems. Other remedies might include ensuring the bankruptcy-remoteness of key assets applied by the infrastructure provider or *ex-ante* measures to ensure that system participants stand ready to contribute financially should business risk crystallize.

3 Governance and regulation of payment and settlement systems

Central banks often now assume a role in the oversight of payment and settlement systems, applying internationally agreed standards to the systems falling within their scope. Their intervention is typically justified by the argument that market participants do not fully consider the consequences of their actions on the rest of the payment system, and, by extension, on the rest of the economy: that is, by arguing that their actions give rise to systemic risk externalities. Many central banks continue to exert influence via ownership of their country's large-value system, or the operation of key components of the infrastructure. In **Part III**, then, we examine the various roles taken by central banks, exploring also the theoretical underpinnings of alternative regulatory and governance structures.

3.1 *Alternative models for governance and regulation*

Payment systems are characterized by increasing returns to scale and strong network externalities. These imply a tendency towards concentrated provision of payment and settlement services, which may not only be a source of systemic risk, but may also have implications for competition, access, pricing and innovation.

One potential mitigant here is mutual ownership, which is common in infrastructure provision. This can go some way towards aligning the incentives of the central provider with those of system participants, ensuring also that the system internalizes a high proportion of the costs associated with the crystallization of risks. However, there may be co-ordination failures among the mutual owners, as well as potential spillovers, beyond the immediate group of members, to the wider economy. There may, then, be a case either for external stakeholder representation in governance or for public intervention.

Issues around competition, access and innovation may be addressed by the competition authorities, if the mandate of the central bank does not extend to ensuring the efficiency of the provision of payment services. But in practice, at least for systemically important systems, there will tend to be close co-operation between the authorities, with the mitigation of systemic risk in payment systems almost always falling to the central bank in its financial-stability role.

3.2 *Central-bank roles in payment systems*

A number of roles might be adopted by central banks to influence risks to financial stability posed by payment and settlement systems. Three principal vehicles are identified in the literature:[6]

- *Public ownership*: an ownership stake in the entity governing the payment system;
- *Operation of the infrastructure*: active engagement in the design, implementation and operation of all, or a subset, of the software, hardware, communication networks, data centres and contingency sites underpinning the system;
- *Oversight of the system*: day-to-day monitoring of system performance to ensure compliance with a set of minimum standards and design principles.

Ownership and oversight, if supported by adequate (perhaps legal) powers of enforcement, can be substitutes. Indeed, a pure oversight model can overcome potential issues around governance failure and user disengagement, which are often associated with public ownership.

[6] Millard and Saporta (2007)

In practice, a variety of models of intervention are applied around the world, reflecting different legacy positions, different political economy environments, and different banking and financial structures. Models of oversight have, however, converged, with an increasing number of central banks assessing their systems against the standards laid out in the *Core Principles for Systemically Important Payment Systems* established by the Committee for Payment and Settlement Systems (CPSS) at the Bank for International Settlements (CPSS, 2001a). In partnership with the International Organisation of Securities Commissions (IOSCO), the CPSS has also established *Recommendations for Securities Settlement Systems* (CPSS, 2001b) and *Recommendations for Central Counterparties* (CPSS, 2004).

4 Future policy challenges for central banks

This convergence in standards for oversight in part reflects the evolution of the infrastructure landscape and, in particular, increasing international interdependencies. In **Part IV**, the final part of this volume, we survey this landscape to identify some of the important emerging themes early in the twenty-first century and examine the regulatory and policy challenges these may pose for central banks.

The payments and settlement landscape is changing in response to the emergence of cheaper technologies, financial innovation, globalization and regulatory change. Developments here pose a number of challenges for central banks and regulators, including setting the appropriate scope for oversight. An important policy question is how to respond to the increasing provision of infrastructure services by commercial banks, particularly where market participants seek multicurrency services. Lessons learned from the financial system stresses of 2007 and 2008 are likely to condition any regulatory and policy responses in this area.

4.1 *Banks operating as infrastructure providers*

Participation in a payment or settlement system may be direct or indirect. Direct participants hold accounts with the settlement agent. Indirect participants, on the other hand, settle across accounts held with direct participants (referred to as settlement banks). Payment systems in which some banks participate indirectly are referred to as 'tiered' systems.

Tiered structures can be liquidity efficient: settlement banks can internalize some of their customers' payments and can also recycle liquidity efficiently by using liquid funds received for one customer bank to settle payments for another. However, tiered structures may at the same time introduce new sources of risk. For instance, indirect participation introduces credit exposures between settlement banks and their customer banks. In addition, a customer bank depends on its settlement bank's ability to raise sufficient liquidity to effect payments on its behalf in real time. Business and operational risk may also be amplified, as the customer's ability to effect payments not only depends on its own operational availability and solvency, but also on that of its settlement bank.

Important questions therefore arise as to how firms providing infrastructure-like services should be regulated. A firm providing settlement-banking services – particularly one that internalizes a significant portion of payments – is, in essence, operating a payment system. Similarly, a custodian bank internalizing securities transactions is, to all intents and purposes, operating a securities settlement system. So, in respect of such segments of their business, should firms offering these services be subject to similar regulatory standards and oversight as traditional payment or settlement systems? More generally, how can the increased systemic importance of these firms be reflected in supervisory standards for banking firms?

4.2 *Globalization, innovation and the changing infrastructure landscape*

One driver of an increasing role for firms offering infrastructure-like services is increased demand for multicurrency settlement services, which might be more readily met by commercial bank providers. To this end, globalization may entail a greater role for firms operating as infrastructure providers. Indeed, the globalization of banking is shaping the evolution of the infrastructural landscape in many ways. This volume, therefore, spends some time considering the various policy issues that arise in this sphere.

The global integration of capital markets has spawned a small number of large internationally active financial groups, operating in many different markets and currencies. Facing a complex web of intraday and end-of-day cross-currency liquidity needs, an increasing number of these financial groups have responded by centralizing their liquidity-management

function. That is, a central Group Treasury takes responsibility for management of subsidiaries' liquidity needs, and co-ordinates the movement of liquidity, and collateral, across parts of the group.

Centralized liquidity management requires arrangements that facilitate the cross-border transfer of securities and recycling of cross-currency liquidity surpluses. There have, therefore, been active calls to central banks to accept a wider range of foreign-currency collateral and to work with private providers to enhance infrastructural arrangements for mobilizing such collateral.

More generally, internationally active banks seek infrastructure providers with international reach. Hence, clearing and settlement systems are establishing cross-border links, and cross-border mergers are becoming more common. Such developments give rise to complex interdependencies between systems and, as infrastructures come ever closer together, these interdependencies might be expected to become stronger.

Financial innovation is another significant source of change in the infrastructure landscape. For instance, OTC derivatives markets have traditionally been cleared and settled bilaterally, typically with significant manual intervention. But as these markets have grown, and back-office capacity has been stretched, new automated infrastructure services have emerged. And, aided by technology, such services are unbundling traditional clearing and settlement functions, introducing competition at narrower niche points in the post-trade value chain. The financial market strains of late 2007 and 2008 have strengthened calls for further automation and centralization of the post-trade infrastructure in this space.

Finally, public intervention – either *ad hoc*, or more sustained regulation or oversight – has historically had implications for the evolution of the financial infrastructure. At the time of writing, the Code of Conduct in the EU has sought to promote competition in clearing and settlement by calling for interoperability and open access between clearing and settlement providers.

And other public-sector initiatives have the potential to alter the shape of the landscape markedly. For instance, the launch of the Continuous Linked Settlement (CLS) system for foreign exchange in 2002 has had implications for the broader infrastructure landscape, both in terms of its creation of new interdependencies and its influence on agents' behaviour.

4.3 *Where are we headed and why does it matter?*

It remains to be seen how the infrastructure landscape will evolve. Whatever the end-game, the changes triggered by globalization, financial innovation and public intervention pose a number of serious questions for central banks and other public authorities:

- What are the implications for system resilience?
- What is the appropriate scope for central bank oversight? And would the conduct or organization of oversight need to change in any way?
- What might be the implications for central banks' operational roles? And might there be implications for the future of central-bank money settlement?