THE HANDBOOK OF BANKING TECHNOLOGY

TIM WALKER
LUCIAN MORRIS

WILEY

The Handbook of Banking Technology

TIM WALKER
LUCIAN MORRIS

WILEY

This edition first published 2021.

Registered office
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom

Library of Congress Cataloging-in-Publication Data is Available

ISBN 978-1-119-32801-8 (hardback) ISBN 978-1-119-32810-0 (ePub)
ISBN 978-1-119-32806-3 (ePDF) ISBN 978-1-119-32809-4 (Obook)

Cover Design: Wiley
Cover Image: © Phaigraphic/Shutterstock

Set in 10/12pt TimesNewRomanMTStd by SPi Global, Chennai, India

10 9 8 7 6 5 4 3 2 1

To Debbie, Tom, Holly and Emily.
– TJW

To Jo, Cameron and Amelie.
– LAM

Contents

Preface

Our goal in writing this book is to bring about a better understanding of the complex world of technology in banking. Of course, there are many books covering just about every aspect of technology, from mainframe computers to cloud computing and from the first programming languages such as COBOL to the latest such as Go. However, books that cover how technology is used in the banking industry are rare to non-existent, which means that those working in the industry have had to rely on documentation supplied by vendors, training courses, the Internet and, of course, by talking with their colleagues in the technology function. Throughout our careers working for many different types of banking and payment businesses, from one of the smallest with a single branch to the largest international banks and various start-ups along the way, we found documentation supplied by vendors, with some notable exceptions, is often incomplete or superficial, only available after attending a training course or only aimed at developers. Attending training courses is often expensive in terms of both money and time and Internet resources can be patchy, of variable quality and sometimes just wrong. One of the reasons we decided to write this book was to fill this gap.

This book is therefore targeted at all readers with an interest in banking and technology, not just technologists. It provides an introduction to the history of banking and an analysis of the current technology landscape that supports the major functions of modern banks, and looks at the opportunities and risks posed by the digital era that we are in. To read this book you do not have to be deeply technical, nor focused only on technology. Instead the book aims to provide all readers with a common understanding of the opportunities and challenges related to technology that all banking businesses must address.

Of course, if you are an executive or manager in the banking industry who wishes to improve your understanding of the technology used in the industry, this book is particularly relevant. It should also provide a good introduction for those who are embarking on a career in technology in the banking industry or who have changed roles in a banking technology function and need to familiarise themselves with the technology used in their new area. It's also useful for those looking to launch a new bank or fintech start-up, even if only to illustrate how technology should not be implemented!

We cover retail and commercial banking, which includes bank accounts, secured and unsecured lending, payments, and payment cards, across all the channels from branches to mobile banking that are in common use, and all of the functions that you'd expect to find in a modern bank, from operations and finance to marketing. We have focused on the technology behind basic bank accounts, payments, and customer interaction and been less detailed about the platforms that support specialised banking products. We don't cover investment banking, investment management, asset management or

insurance, which would probably require books in their own right. We have drawn on our experience in the banking sector in Europe, the Middle East, and North America, although readers will quickly understand that we tend to draw most deeply on our experience in the UK, where we have spent most of our careers.

Of course, banking businesses are often complex and the technology they use is also complex, wide-ranging and anything from five or more decades old to right up to date. Our aim is to provide an overview of the technology in use so that if you are a manager or executive outside the technology function you should be able to have more informed conversations with your colleagues in the technology function, and if you are in the technology function you should better understand the wider context of the work you are doing or managing. If you need to get more technical, you can consult the references and further reading lists that we provide.

As we are covering technology in the banking industry, we often tend to use the term *bank* when in reality what we have written applies to businesses that provide banking services and which may not, strictly speaking, be banks. If we were to follow the UK's banking regulators, an organisation can only be called a bank if it is licensed to take deposits. For example, credit card issuers and lenders that do not take deposits are not banks but they do provide banking services, and many aspects of what we have written in this book are also relevant to them. Given that payments are integral to banking, we have devoted a chapter each to card payments and to interbank payments.

Acknowledgements

We started thinking about and then writing this book several years ago. However, progress was patchy as work and family life (including Lucian's cycling accidents and the death of Tim's teenage son from leukaemia) took priority. By the start of 2020 we had written less than a third of the final manuscript. However, we decided to take advantage of the lockdown due to the Covid-19 pandemic to focus on completing the book. Our families sustained us while we beavered away, for which we say a huge thank you.

Although we worked together previously, this exercise has taught us that co-authoring a book can be challenging but is ultimately rewarding. Reviewing each other's writing made us realise we had different writing styles and points of view on various topics, and that sometimes what one or the other of us had written was just plain wrong. At times we have even been a little short with each other. However, we were always able to discuss and resolve these differences with equanimity while also still striving for quality and completeness. If there are mistakes in this book, we accept joint responsibility.

Finally, the last two chapters are relatively opinionated. The reader is perfectly entitled to disagree with what we have written.

CHAPTER 1

Introduction

1.1 Banking and the Rise of Technology

The banking industry, in many varied forms, has verifiably been in existence for at least four millennia. Beginning as a simple money and commodity management activity that supported early merchants and royalty, banking has gradually evolved into today's model: a complex, highly connected network of businesses that spans the globe. While it can appear that the development of the banking industry has generally taken place at a rather sedate pace, in reality banks have been established, grown and consolidated into ever larger organisations seemingly non-stop for centuries.

The advent of computers and, later, the Internet have had a dramatic impact on how banking has been conducted in recent times and ended the long-standing trend of opening more and more branches across the globe. Banks have changed their products and services, developed credit and debit cards, introduced computers to improve efficiency, built multi-channel digital banking platforms and, in many geographies, significantly reduced their high street footprint. They are moving from being brick-and-mortar businesses to what seems often to be purely digital utilities. However, the emergence of digital-only propositions may allow new businesses to develop with a much lower cost to serve, better products and services and more convenient customer access. A *digital-first* world of banking, it is argued, would not only replace the legacy brick-and-mortar world, but has the potential to deliver business models that will out-compete the incumbents, businesses that have grown huge, cumbersome and complacent. Visionaries and digital advocates in the financial services market argue that incumbent banks are, like the dinosaurs, plodding slowly but surely towards their inexorable extinction. On the other hand, large incumbents have the advantage of huge economies of scale and existing customer bases, meaning they pay less for deposits and make more money from each of their existing customers. Although there is evidence to show that the costs of complex, legacy technology are higher than modern technology, the massive scale of incumbent banks may mean that this cost can be borne for long enough for these banks to modernise their technology estates. Thus, the stage is set for a struggle between the old and the new.

The outcome of this struggle is far from a foregone conclusion and, in fact, the history of banks and banking institutions implies that the bigger ones eventually acquire the smaller ones – for more than a century the total number of banks in developed markets has been declining and this trend continues even up to the present.

Some observers have predicted that the shift to digital will sideline incumbent organisations in a different way, akin to what happened in the mobile telecommunications industry, where huge and highly profitable new businesses offering a new paradigm – the smartphone – were built. The mobile network providers no longer have access to the bulk of the revenue in this industry and are purely utilities with varying levels of profitability and return on capital in different markets around the world. Could incumbent banks just become utilities providing vanilla banking products with the customer relationships intermediated by new organisations with better technology? In practice, many digital new entrants are offering just the same banking products as incumbents, with a better mobile app, and this doesn't feel like a paradigm shift like the smartphone was. Also, it appears likely that only some incumbents will manufacture products for others and the rest will resist being intermediated unless forced by regulators, such as in the EU, where the Payment Services Directive 2 (PSD2) has resulted in banks having to offer an open banking interface for use by third-party service providers.

There is a third point of view: that to even present the current situation as incumbents versus new entrants and fintech start-ups, or old versus new, may not even be an appropriate or accurate representation of the true picture.

Where many analysts and commentators originally expected conflict (as early as 1994 Bill Gates famously made his big bank dinosaur speech),[1] it is now becoming clear that the relationship between new and old will be much more nuanced, with start-ups as likely to cooperate with incumbents as they are to compete.

So, we believe that the predictions of apocalypse for the incumbents within the banking industry are far from certain. While the digital agenda is creating opportunities for new entrants and does pose a risk to incumbent organisations, the rise of innovative fintech solutions and business models also provides many opportunities for established entities. There is no doubt that a significant amount of work, both forward-looking and in remediation, must be carried out in order for the incumbents to remain competitive and position them to seize these new opportunities, but their size and scale offers stability, established market access, lower unit costs and the income to fund the required changes – all key attributes that the new entrants and fintech start-ups typically lack. Over the last seven decades the established banks have proven themselves more than capable of adjusting to, and adopting, new technologies (computers, ATMs and Internet banking, to name a few) and any suggestion that a new wave of technologies means certain destruction for them may therefore be premature or even wrong. Not only do they have a proven history of adaptation, but many also have the balance sheets to buy themselves out of trouble should they need to, either through investing in joint ventures with up-and-coming start-ups or through outright acquisition of potential competitors.

Incumbent organisations may well struggle to adapt and there will inevitably be some organisations that will fall by the roadside (which usually means being acquired), but any significant change in market conditions typically results in casualties among the established order in that market. To survive and prosper, the leadership of a bank needs to recognise that it is a fundamentally digital business and that having a sufficient understanding of technology, its uses and sources of competitive advantage is essential.

1.2 The Challenges of Technology in Large Banks

There are several typical challenges that a bank's leadership faces relating to technology. For example, the leadership will have to decide whether its bank can persist in using legacy banking platforms. In order to meet the demands of the new digital economy our incumbent banking organisations require substantial changes within their organisations, and not just in the technology itself. Bank leadership teams need to acknowledge that the banking business is essentially becoming one of technology. Banks are now, at their core, digital businesses, and the leadership team must accept that and ensure that it has the skills at the top of the bank to manage a technology organisation. One of the reasons we wrote this book is to help to raise the level of understanding of technology among bank management.

Many of the platforms and systems that sit at the core of banking businesses are legacy environments, containing dated software and hardware, that have been built up over many years. Stories continue to circulate about the age of some of these systems. For example, the UK newspaper *The Telegraph*, as recently as December 2016, published an article claiming that some banking platforms in the UK still run on pounds, shillings and pence,[2] the UK's currency before decimalisation in 1971. We do not know whether this story is really true or not, but it was certainly an urban myth that occasionally surfaced in our conversations across the industry during our careers. While the latest developments may have been carried out in modern software development languages, the core of these platforms is still legacy and is often complex and poorly understood and therefore prone to going wrong when changes are made. Not only are these platforms getting on in age, but so are many of the developers who understand them. Such systems are often blamed for the seemingly high cost base of technology, the shortcomings of various products offered by the bank, and the lack of flexibility to introduce new products and services. On the other hand, re-platforming is expensive and full of risk – we are aware of several banks around the world that spent huge sums replacing their core banking platforms, and some, such as TSB in the UK,[3] that ran into highly public difficulties.

In parallel with maintaining legacy banking platforms, it would appear that many banks have struggled to maintain appropriate controls over the ongoing development of their technology estates. Of course, there are many reasons for the huge diversity in technology estates, which include business-led decision-making with technology standardisation low on the list of priorities, ongoing developments in technology with consequent dead ends and obsolescence, and accumulation of technology variety through mergers and acquisitions. Consequently, modern banking technology estates are complex and often poorly understood, even by the technology functions that run them. Some estates are so large that even tracking the hardware and software within them can be a significant task and many large modern banks struggle to do even this effectively, let alone calculate the input costs for each service that they supply and for which they charge the business units that consume them. Of course, there are usually no easy solutions to the problems posed by this complexity and we are often reminded of the supposedly Irish response to a request by a lost traveller about how to get somewhere which goes *I wouldn't start from here if I were you*.

Banks' technology functions can be as complex and politically charged as any other function in a large business. In large international banks the internal technology organisations are often huge, containing tens of thousands of workers distributed across dozens of sites, and are dependent upon a diverse supplier ecosystem. The structure of these organisations varies from bank to bank and over any given period, often oscillating between business-aligned and technology-driven structures (as discussed in Chapter 9). When business alignment is the primary driver, individual technology functions align with business units and while this may provide greater control to the business units, it can also result in technology functions that associate more with their business unit than with the CIO's office. This can lead to political in-fighting and huge inefficiencies, such as duplication of roles, technologies and third-party relationships within the broader organisation as similar capabilities are established and technologies built to support the same general need. The alternative model, with alignment to technical horizontals, can deliver much greater technological competence, reduced potential for business-driven in-fighting and lower costs, but does so at the risk of alienating the business units when they cannot get the service they believe they need to support the ongoing running and development of their business. When this happens, we have seen business units develop their own in-house technology functions, resulting in additional technology costs, duplication of roles and systems and a lack of consistency across the technology estate, in much the same way as the business-aligned model.

On top of the challenges associated with managing an international technology organisation, the current low public regard for banking,[4] namely that it is a staid industry, full of untrustworthy middle-aged men in suits with outdated views on technology,[5] makes it harder to recruit graduates and technology professionals that can bring the much needed skills and mindset the industry needs to fully grasp the opportunities presented by the digital world. Also, the ambitious perceive they can have more impact and make more money elsewhere. The talent that is clearly needed in technology departments within incumbent banks is being drawn to the more dynamic environments of technology-centric companies such as Apple, Facebook and Google and the fintech companies that the incumbent banks fear.

Finally, the scale of technology functions, and the complexity and scale of the technology itself, present challenges for the processes used to manage and govern them. These processes often represent legacy practices that are engrained in the culture and mindset of the technology function. Changing these established ways of working requires a supreme effort and a determined executive team. Despite recent high-profile technology failures, or perhaps even because of them, internal technology functions are often conservative by nature and married to legacy processes and ways of working that are proven, despite being inefficient. In order to adjust to the threats and opportunities posed by the digital agenda, the technology functions will first need to address their legacy operating models and adjust to new ways of working.

1.3 Navigating This Book

In Chapter 2 we provide a review of the history of banking and a snapshot of the current state of the banking industry. This chapter is not essential reading, but it provides context for the rest of the book.

Chapter 3 provides a technology primer, building up a picture of the technology required to support a simple bank that provides bank accounts, credit cards and an Internet banking channel. It includes an overview of core banking platforms, databases, mainframe technology (which is still prevalent in large banks), how platforms are made highly available and how different platforms are integrated, including both service-oriented architecture and microservices. It finishes with a review of data analysis, touching on data warehouses, data marts and data lakes. Less technical readers may not wish to read all the sections in this chapter.

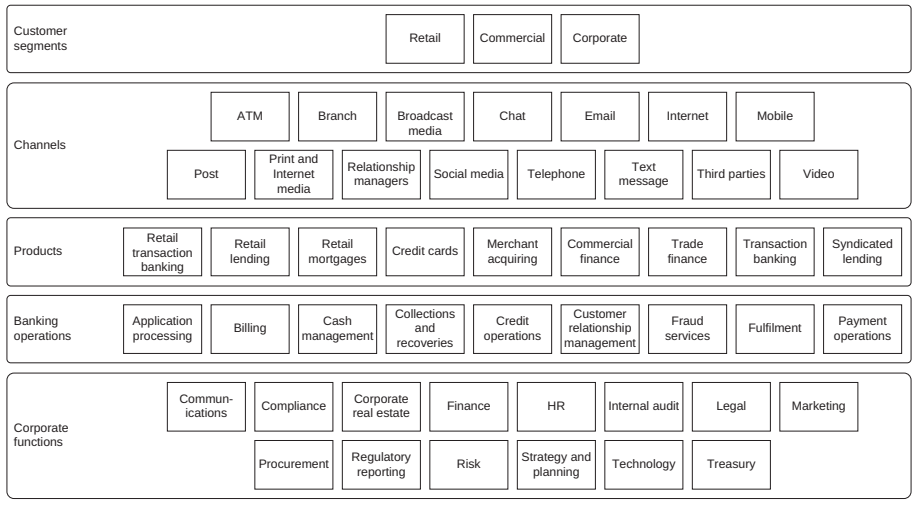
The remaining chapters are sequenced broadly according to the model of a bank we show in Figure 1.1. At the top, it shows the customer segments the bank serves – in this case retail customers, commercial customers (i.e. small and medium-sized enterprises) and corporate customers (i.e. large enterprises). Of course, some banks may use finer segmentation such as underbanked, mass retail and premier retail in place of a single retail segment. The largest banks are likely to have a wholesale segment, which includes the very largest corporations and other banks as their clients. Normally, each segment will have a corresponding business unit inside the bank with a member of the executive team heading it up and marketing, product management, relationship management and sales staff in the unit.

The customers and potential customers of the bank engage with it through various channels, as shown in alphabetical order in the second layer of Figure 1.1. Note that these channels may technically overlap – for example, emails are usually sent over the Internet, but both are regarded as distinct channels by our bank. Similarly, telephone calls may be made using standard telephone lines, from mobile phones or over the Internet. However, every bank needs to consider if and how it will deal with such a channel. Some channels will have corresponding business units, such as the chat, email, telephone and video channels which would be handled by the bank's contact centre(s). We cover the Internet channel in Chapter 3 and then the other channels in Chapter 4.

The business units that correspond to the first two layers of this model are occasionally called the *front office* of the bank.

The categories of products that the bank offers are shown in the third layer down in Figure 1.1. These are ordered so that the products with more relevance to commercial and corporate customers are to the right-hand side. Retail transaction banking covers retail current accounts and deposit accounts and retail lending encompasses both unsecured lending and lending secured on assets other than property (e.g. vehicle loans). Commercial finance includes asset-based lending, asset finance, leasing and sales finance such as invoice factoring and discounting. Transaction banking is the provision of bank accounts and services such as cash management to commercial and corporate customers. Syndicated lending is the participation in and organisation of loans underwritten by multiple lenders to corporate customers. For each product type there will sometimes be a corresponding business unit within the bank, particularly for credit cards (which, along with other payment cards, we cover in Chapter 6) and in some cases the product type will be owned and managed by the corresponding business segment. Many banks offer insurance products to their customers, but we do not cover these in this book. Some banks also offer specialist products based on derivatives to corporate clients to manage interest rate and foreign exchange rate risks, which we would class as investment banking products and so we do not cover them.

FIGURE 1.1 Operating model of a bank.



The products typically draw on some of the banking operations services shown alphabetically in the fourth layer of Figure 1.1, which are sometimes said to comprise the *middle office*. For example, application processing covers the processing of applications for products by new and existing customers, billing covers the billing of customers (typically more complex for larger corporate customers) and credit operations covers deciding whether to lend to a customer and at what interest rate. We cover these services in Chapter 5.

Finally, the lowest layer in Figure 1.1 shows the range of corporate functions a bank may have, i.e. the internal organisational units of the bank (apart from the customer-facing business units aligned with the customer segments in the first layer of the model). Many of these would be described as being in the *back office*, although this varies from bank to bank. In larger banks all the functions shown may exist, whereas in smaller banks some may be combined. For example, communications and marketing are often a single team and finance often includes regulatory reporting and treasury. We cover these functions in Chapter 8, apart from technology, which is the subject of Chapter 9.

Note that, although we have used the classification of front, middle and back office in this section, there are no standard definitions of what is included in each category and many banks do not use this terminology. Some banks use the classification of distribution (broadly equivalent to the sales functions in the front office plus marketing), manufacturing or production (covering the various products manufactured or produced by the bank and the servicing of them) and corporate services (equivalent to the back office). We will tend to avoid these classifications in the rest of the book.

1.4 References

1. Zack, J. (2015). Is Microsoft's Bill Gates right when he says banks are technology? *American Banker* 160 (5): 11A. https://www.americanbanker.com/news/is-microsofts-bill-gates-right-when-he-says-banks-are-technology (accessed 22 July 2020).
2. Evans, R. (2016). Some banks' systems still work in pounds, shillings and pence. That's a terrifying thought. *The Telegraph* (12 December). https://www.telegraph.co.uk/personal-banking/current-accounts/banks-systems-still-work-pounds-shillings-pence-terrifying-thought/ (accessed 2 June 2020).
3. Morrison, C. (2019). TSB IT meltdown cost bank £330m and 80,000 customers. *The Independent* (1 February). https://www.independent.co.uk/news/business/news/tsb-it-failure-cost-compensation-customers-switch-current-account-a8757821.html (accessed 22 July 2020).
4. White, L. (2018). British public don't trust banks 10 years after crisis, survey finds. *Reuters Business News* (16 August). https://uk.reuters.com/article/uk-britain-banks/british-public-dont-trust-banks-10-years-after-crisis-survey-finds-idUKKBN1L11EL (accessed 22 July 2020).
5. Hodge, N. (2016). U.K. boardrooms still 'pale, male, and stale'. *Compliance Week* (7 September). https://www.complianceweek.com/uk-boardrooms-still-pale-male-and-stale/2938.article (accessed 22 July 2020).

CHAPTER 2

The History and Current State of Banking

This chapter provides a brief introduction to the history of banking, the origins of cash and gold, and the importance of the branch in banking, before investigating the drive for scale that has been a defining element of the industry over the last century or so. We then look at the development of modern banking products and services and explore how technology has come to play such an important role in modern-day banking. The chapter ends with a review of the current state of banking.

When researching this chapter, it became apparent to us that what is accepted as historical fact is sometimes not correct, and sometimes there are conflicting or unsubstantiated claims for being the first to have done something. We have therefore provided extensive references in this chapter and in some cases we have qualified what we have written. There remains the risk that we have missed the actual first instance of a banking product or service because it isn't documented or because there were no references to it in the materials we consulted.

2.1 A Brief History of Banking

Much as it would be amusing to claim that banking is the earliest profession, this is, alas, untrue. One could argue, though, that without money, and therefore banking, it would be hard for any other walk of life to claim to be truly professional. The origins of banking can be traced back several millennia. From almost 2000 BCE there is evidence that the Babylonian temples provided a range of banking services, with their records kept on clay tablets.[1] Initially, as in other early cultures, this took the form of safeguarding of deposits of grain and then other agricultural products and ultimately precious metals.[2] Obviously, the receipt issued for a deposit had a value, which could be exchanged for other goods, and over time it also became possible to transfer the ownership of a deposit to another person. Another development was of temples making loans, initially of seed grain to be repaid after the harvest and then of money in the form of specific weights of metals such as silver, with interest payable by the debtor. Banking was important enough by this point in mankind's development that in the eighteenth century BCE, Hammurabi, the ruler of Babylon from around 1792 BCE, codified laws to govern banking practices such

as the advancement of capital and the rate of interest to be paid on loans.[3] These laws were included in the Code of Hammurabi, the most famous copy of which was engraved on a stone slab (or stele) that was subsequently found in Iran in 1901 and is now in the Louvre Museum in Paris. So, in summary, the Babylonians developed three core aspects of banks: safekeeping of items of value, lending, and transfer of ownership of deposits (i.e. payments).

Mediterranean civilisations developed those early banking practices through the following 2,000 years, with the maritime nature of the Greek civilisation having a significant impact upon the development of moneylending. By the fourth century BCE, the civilisation of the ancient Greeks saw private individuals, temples and public bodies taking deposits, making loans, settling accounts (i.e. transferring the ownership of deposits) and changing money.[4] Changing money was required because by this period the Greek city states were separately issuing coins, and so travelling traders needed to be able to exchange coins from different city states. Maritime loans enabled traders to pay for cargoes, with the loan secured on the trader's ship, and the debt being cancelled if a trader's ship sank before returning home. The rate of interest charged was up to 30%, commensurate with the level of risk.[5]

In Egypt during the Ptolemaic era (305 BCE to 30 BCE), royal grain depositories were developed into royal banks that operated the first known giro system.[6] The basic concept of a giro system is that each customer of a bank has an account into which he or she makes deposits (often of grain, in this case) and that the bank records. Internally, the bank consolidates all the deposits and keeps them in one large store rather than keeping each customer's deposits separate.[7] A customer can then request that the bank transfer some of their deposits to another customer, and the bank updates its account records accordingly – without any physical transfer of the deposits. In particular, Egyptians could pay taxes to the government that were due on their produce through this procedure and the government made payment to its officials in the same way. This procedure is the equivalent of today's credit transfer. Records from the depositories and banks were sent to the equivalent of a central bank in the capital, Alexandria. In addition, a customer with deposits could issue a cheque to a third party, who could claim payment at any state bank.[8] Note how the cheque process differs from the giro or credit transfer process – the beneficiary presents a cheque to the bank and can receive payment in cash or equivalent, whereas the payer instructs the bank directly with a giro or credit transfer and payment is made into the beneficiary's account. It is the ability to use a network of banks with payments possible at different branches that marks out the Egyptian banking system of this era.

Over the following centuries, banking continued to grow in importance and eventually became such a key element of commerce that under the Romans it became regularised and various banking roles were recognised (and at times even appointed) by the state. Roman law even set out that bankers had unlimited liability for funds deposited with them (a principle that we will see later in eighteenth-century England) and how they should record transactions.[9] However, it is also recognised that the Romans did not develop a centralised network of banks and giro payments comparable with the Egyptian network.[10]

Of course, banking was not unique to the Mediterranean world. There is evidence from India during the early Vedic period (between 1500 BCE and 1200 BCE),[11,12] and from China in the third century BCE, of established banking practices.[13] In fact, it may well

be that these cultures have a longer continuous history of banking practices than does the West.

There is little documentation of the history of banking in Europe after the fall of the Roman Empire until the eleventh century, and various authors record (see, for example, reference 2) that the main form of banking was money changing, given the issuing of different coins by many kingdoms. The influence of Christianity had a major impact on early banking – usury (originally defined as charging any interest on a loan) became frowned upon by Christian culture, banned for all clergy by the first council of Nicaea in 325 and was eventually prohibited altogether by Charlemagne in the ninth century.[14] Fortunately for the Western world, usury was not banned in all cultures. The Jewish faith, in common with Christianity and Islam, bans usury. However, Jewish law interprets the relevant verses in the Torah (Exodus 22:24, Leviticus 25:36–37 and Deuteronomy 23:20–21*) and other books in the Bible as specifically forbidding charging interest on loans to fellow Jews but allowing it to be charged on loans to non-Jews.† With moneylending being one of the few careers left open to them in Europe, many Jews entered the business, but at a cost of significant prejudice and jealousy from the majority non-Jewish population, who resented their wealth. This prejudice, and the ruthlessness of royalty and the Church, who coveted their wealth, resulted in pogroms that led to confiscation and theft of their wealth and the horrific treatment of Jews in general.[15]

During the twelfth century, the Crusades created demand for the development of a system of letters of credit, operated by the Knights Templar. A pilgrim or crusader could deposit money at a temple in their home city and receive a letter of credit in return, which they could exchange en route or at their destination.[16] The Templars also accepted valuable items for safe-keeping and used them as collateral for loans of money (which they had sourced from donations).

While it is true that in earlier periods (despite usury being frowned upon) many Christians, including the Church, had participated in or supported lending, in the thirteenth century Christians started to become more openly involved in banking. In Lombardy, the new breed of banker used double-entry bookkeeping,‡ one of the backbones of modern accountancy.[17] This was the start of the growth of Italian banking. As the Italian merchant states grew in importance, so did the importance of banking to the merchant economy. In the fifteenth century, the Medici family established its famous banking dynasty, and in the sixteenth century, the first modern state-owned bank, Banco della Piazza di Rialto, was established in the Italian city-state of Venice to hold merchants' funds safely, more so than some private banks which had failed prior to its establishment.[18] Like the private banks before it, it enabled transactions to be made between merchants without the physical exchange of coins. The establishment

*The precise verse numbers may differ depending on the version and/or translation of the Torah.

†Given the Christian Bible contains the same verses, there was also a view that Christians could lend to non-Christians. Given there were many more Christians than Jews in the Middle Ages in Europe, Jewish lenders to Christians had a much bigger market than the other way round. Islam, on the other hand, prohibits all usury, even by those of other faiths.

‡We have seen claims that forms of double-entry bookkeeping may have been used by the Romans and the Muslim Empire before the Italian bankers, but these appear to be based on the fact that the Italian bankers drew on the teachings of the Muslim Empire via Jewish traders and that all of the basic technologies required for accounting (e.g. arithmetic, commerce, credit, capital, etc.) were in place in the Muslim Empire; see reference 17.

of state-owned banks was later followed by the creation of national banks, with the first example being the Bank of Sweden (or Riksbank) in 1668, followed by the Bank of England in 1694 and the Bank of Scotland in 1695. Governments created national banks for a variety of reasons. The Riksbank was instituted to provide commercial banking services (the safekeeping of deposits and lending) following the collapse of the first chartered Swedish bank, the privately owned Stockholms Banco, which was effectively controlled by the Swedish king. It had collapsed after issuing loan notes (including to the king) greater in value than the deposits it held, which were exhausted when recipients of the loan notes cashed them in.[19] With parallels in modern central banks being granted autonomy, the Swedish parliament, the Riksdag, had learnt from the collapse of Stockholms Banco and kept control of the Riksbank away from the king. The Bank of England was created to provide funds to the government to pay for its involvement in the Nine Years' War with France.[20] The Bank of Scotland was established as a commercial bank offering banking services to Scottish businesses and was originally prohibited from lending to the Scottish government. Within a year of being founded, the Bank of Scotland was issuing banknotes (which reduced the need to use heavy coins).[21] National banks have since often gone on to play fundamental roles in the banking industries of their respective nations, becoming central banks typically with remits to manage monetary policy (including controlling the issuing of banknotes), to provide settlement accounts for payment schemes and to regulate the banking industry.

2.2 Cash, Gold and Digital Money

While banking has existed since at least the eighteenth century BCE, the concept of money in the form of coins is a much later development. Prior to the development of coins, any manner of items was used as a means of stored value. Such commodity money[22] could range from pieces of precious metal including gold, lead, silver and tin, and stones (precious and otherwise) to animal teeth, cattle, feathers, fur, rice, salt, shells and other materials.[23] In ancient times, in the Middle East, specified weights of precious materials (mostly silver) were used as the basis of money. One such weight was referred to as a shekel, a term that became associated with the coinage that was developed. The shekel is probably the world's oldest term for a unit of value that is still in use today.[24]

The British Museum has in its collection some of the earliest known minted coins, dated around 550 BCE and which originate from the Kingdom of Lydia, in what is now western Turkey. A coin is typically defined as a metal object with a fixed weight and size (and therefore density and composition) for its value, and that is stamped to give it official standing. It is believed that the Lydians invented coinage about a hundred years earlier[25] and were possibly the first to do so because they had ample sources of gold in the form of electrum, a naturally occurring amalgam with silver.[26] It is also believed that coins were invented in China around the same time.[27] Following this, at various times, thought to be no earlier than 650 BCE and no later than 500 BCE, most of the Greek city-states started issuing their own coins.[28] As the first Lydian coins were made from electrum, in which the relative proportions of the constituents could vary, two coins of the same size and shape may not have weighed the same, casting doubt over their value. However, over time the Lydians found a method of making pure gold coins which meant that two coins of the same size and shape weighed the same, and so purity could be assured. The Lydians made

coins in a range of sizes that adhered to strict weight criteria, which enabled a specific value to be allocated to each size of coin. This ability to maintain a specific value for any given coin is fundamental to the whole basis of coins.

The creation of a system of coinage is a significant step forward. It creates an easily portable and recognised source of value that can be traded for all manner of goods and services. While coins have always suffered from a risk of debasement, such as through forgery or coin clipping, from less reputable members of society, they have proven to be a remarkably resilient means of storing value since they were first created.

The key to the success of coinage (and, later, also banknotes) has been in its ability to maintain a value. The basic premise of stored value in coins throughout history typically came down to one of two approaches.

First, and originally, coins had intrinsic value, as they were made from precious metals. There is no dependence on the issuing authority to be able to refund the bearer of the coin and its value is arguably very stable (though, of course, if the value of the constituent metal or metals changes then the intrinsic value of the coin may change as well). While in practice it appears that there were many cases where the face value of the coin was actually slightly more than that of its constituent parts, this did not seem to matter as long as the participants in the market agreed that the face value was broadly representative of the coin's actual worth. A risk with this approach is that the metal used for coinage can become scarce as the economy develops and people are then forced to resort to other ways of exchanging value (such as promissory notes and receipts for deposits at banks).

Second, and as is seemingly universally the case today, coins had token-based value – where the value of a coin is different (typically greater) than the metal(s) it is made from and is based upon trust in the issuing authority. In this case, the issuing authority is held to account for that value and therefore is, in essence, issuing a promissory note. This is, of course, also the case with banknotes and is the origin of the promise to pay the bearer statements found on many modern notes. In theory, the holder or bearer of the currency can turn up at the issuing authority, hand in the coin or banknote, and be reimbursed the value in gold.

The second of these two approaches, which is termed representative money,[29] essentially requires the issuing authority to hold a reserve of gold to cover the value of the coins and banknotes it issues. Without this reserve, or at least the trust that the means to pay exists, the cash loses its value. From the eighteenth century onwards the use of a gold standard became widespread, with various nations implementing laws to enforce the direct relationship of their currency to their gold reserves. However, the ability to do this is obviously directly related to a country's ability to hold reserves and this was always an issue during periods of stress – civil wars, world wars and the depression of the late 1920s and 1930s all put pressure on the ability of many countries to maintain a gold standard as governments used up their reserves to pay for their armed forces or to fund social programmes. France and Germany suspended their gold standards in 1914 (although France returned to it in 1928, with the franc at a fifth of its value compared to when convertibility was suspended, and then left it in 1936), the UK abandoned it in 1931, followed by many countries in 1932, and then the US in effect suspended convertibility in 1933 by stopping banks issuing gold, preventing gold exports, making it illegal for people to hold gold bullion, getting the Federal Reserve to issue notes that were not backed by gold and implementing a policy of devaluing the dollar in terms of

the gold equivalent.[30] In each case these governments worked on the basis of issuing fiat money – money that has value because the issuer says that it does.[31]

At the end of World War II, major western currencies agreed to adopt a gold exchange standard* with fixed exchange rates with the US dollar which was convertible to gold on demand by the participating countries. This arrangement lasted until 1971. The beginning of the end was arguably a decision by the French government in 1965 to convert its holding of US dollars to gold. By the start of the 1970s, the ability of the US to honour the agreement was under pressure due to the challenges the US had from funding the Vietnam War, high US unemployment and high US inflation, all of which were exacerbated by a balance of trade that moved from surplus to deficit.[32] This meant that the US had printed more dollars than it could convert to gold. West Germany withdrew from the standard in May 1971, after which other participants swiftly took steps to redeem their dollar reserves. The US effectively ended the agreement in August 1971 when President Nixon suspended convertibility of the dollar to gold. Over the course of the twentieth century there was a general move away from the use of gold standards and no countries were using them by the end of the century.

Today's position is therefore an interesting one in that we now live in a world of fiat currency – where currency is legal tender because the issuing nation declares it so. In practice, the issuing authority typically does not hold enough gold (and does not have to) to cover the value of the coins and banknotes that it has issued. In most cases this system runs well enough, but in periods of stress some struggling governments have been known to resort to money printing to address their immediate issues. While the printing of money may have appeal in addressing some short-term issues, it inevitably results in longer-term woes for the issuing authority. For example, in Germany in 1923,[33] in Uganda under Idi Amin in the 1970s,[34] in Zimbabwe under Robert Mugabe in 2007–2008[35] and in Venezuela under Nicolás Maduro in 2017–2019,[36] this policy led to massive hyperinflation. When this occurs, it damages the economy and renders savings worthless, often leading to a flight to more reliable currencies.

In the modern era the use of fiat currency has seen a significant move away from hard currency (coins and banknotes) towards digital money. In most modern economies this and modern banking methods mean that the amount of money in the economy is significantly higher than the actual value of minted currency and our dependence on coins and banknotes in support of our daily spending has been reduced significantly. Consider a customer taking out a loan with a bank. The act of the bank's setting up the loan is to set up a loan account on its banking platform (or in a paper ledger, before the days of computers) with a negative balance for the amount of the loan, and deposit the same amount into the account of the customer (i.e. add it to the account's balance recorded on the banking platform or in a paper ledger). In effect, by doing this the bank

*With a gold standard, a member of the public can exchange a banknote for the equivalent in gold, with the rate of exchange (the weight of gold received per unit of currency) normally fixed or at least stable. With a gold exchange standard, a group of countries agree to fix the exchange rates of their currencies relative to the currency of one of the group (the reserve currency). The reserve currency has a fixed rate of exchange to gold, so that if one country collects a large amount of the reserve currency, it can exchange it on demand for a known amount of gold. The Bretton Woods agreement of 1944 put in place a gold exchange standard based on the US dollar as the reserve currency with $1 worth 1/35 of a troy ounce of gold.

has created some money. If the customer uses the loan to pay another customer of the bank – a seller of goods, say – by transferring some money from their account to the account of the seller, no physical cash is required. In fact, if the bank's customers only pay each other by transferring money between their accounts at that bank, the bank never has to worry about having any assets to back up the money it created.* However, in reality, more often than not a customer of the bank will pay a seller who has an account at a different bank by transferring funds to the seller's account at the other bank (paying the seller using a debit or credit card ultimately has the same outcome). At this point, the first bank must transfer some assets to the second bank. In a modern payment system, these would be assets owned by the bank that the central bank or some other trusted party holds (typically called reserves), which could include gold but may just be money deposited by the bank at the central bank. Such reserves are of course finite, so a bank could run out of reserves if it issues too many loans. Banking regulators set limits on the amount of lending banks can do relative to their reserves to prevent this happening. Note that as a bank gets larger and increases its share of the banking market (e.g. through faster growth than its competitors, acquisition, or merger), on average a larger proportion of the money it creates stays with it. Also, there are typically millions of small interbank transfers a day, so to settle these payments only net amounts are transferred between the banks. Where electronic payments dominate, the value of digital money will outweigh the value of banknotes and coins in circulation, and this is indeed the case according to central bank data – in the UK only 3% of money is in the form of cash.[37] Also, banks do not have to cover all their deposits with reserves, but they do have to have enough reserves to cover their likely daily outflow of funds plus some extra for unusual levels of outflow. In the UK, 18% of money is held by the banks as reserves at the Bank of England and 79% as customer deposits at the banks.

While most countries around the world, with the obvious exception of those in the eurozone, still maintain their own currency we can now imagine a future where money becomes completely digital. We are now approaching a point where virtual currencies (Bitcoin, for example, which has no underlying guarantor of value) may offer potential alternatives to national currencies. While initial attempts to create virtual currencies have not so far been wholly successful, the ongoing development of technology, particularly concepts such as the blockchain and, even more importantly, cybersecurity (cryptocurrency exchanges, which facilitate the buying and selling of cryptocurrencies such as Bitcoin, have regularly been attacked by hackers[38]) are making them more viable. We return to virtual currency in Chapter 10.

2.3 Branch Centrism

The historical link between banking and physical currency meant that before the arrival of technology most banking products and services could only be provided through physical interactions. Where physical currency was in common usage, there was a natural requirement for bankers to have a physical presence in the market to provide funds. As noted at the start of this chapter, early forms of banking were often closely associated

*Note that this was equally true before banks used computers.

with temples, and even the Bible makes reference to money changers in the temples of Israel (Matthew 21:12–17, Mark 11:15–19, Luke 19:45–48 and John 2:13–16), but it was not long before bankers started to establish a separate presence within the marketplace itself. Due to the nature of trade, it was a natural logical extension to establish networks that could provide services at the ends of trade routes.

However, in Europe it was only in the later Middle Ages that banks established branch networks. The reliance on a physical presence as the only channel through which banks could engage with their customers dictated that in order to grow their businesses, until the twenty-first century, the banks' primary routes were therefore through the geographical expansion of their branch networks and the acquisition of other banks.

For a variety of reasons, including regulatory, cultural and geographic, banking has developed differently around the globe and this is reflected in the development of branch networks and the differing perceptions on the need for, and reach of, these networks. In some countries, banks limit themselves geographically. In Germany, for example, the regional banks or Landesbanken are publicly owned banks that originally funded local projects and provided wholesale banking services to the local savings banks called Sparkassen (legally and financially independent organisations that do not have owners or shareholders[39] focused on a specific geographical area) inside their respective regions.* There are also member-owned co-operative banks, Volksbanken and Raiffeisenbanken, also normally focused on a specific urban or rural geographical area.

Like Germany, Austria has a system of nearly 500 local Raiffeisenbanken that typically own regional Raiffeisenlandesbanken. It also has systems of local Sparkassen and Volksbanken, although these have not fared well. The Sparkassen saw various waves of consolidation from 124 in 1989 to 74 in 1994 and then 46 by 2016, with the majority becoming listed entities rather than member owned.[40] The Volksbanken came under pressure following the banking crisis in 2008, with 51 Volksbanken consolidated into 10 by 2015.[41]

In Spain too, the concept of local banking was embedded in the cajas (savings banks) and cajas rurales (rural savings banks), both of which were established in the nineteenth century. However, the geographical restrictions on the cajas were lifted in 1988[42] and many of them expanded rapidly outside their historical areas and diversified. After the banking crisis in 2008 a large number of the cajas failed, with local political interference in their management and operations having caused them to overextend and make risky loans suggested as a contributory factor.[43] The number of cajas dropped from 45 in 2007[44] to just two (Caixa Ontinyent and Colonya Caixa Pollença) by 2020, with the rest consolidated into various listed banking groups or bought by other banks.[45] In 2020, there were still 29 cajas rurales, down from 77 in the 1980s, all of whom were by that point members of a federated group that provides shared services to its members.[46]

*The Landesbanken were not formally geographically constrained and in the 1970s several developed their businesses internationally. After the EU banned the public guarantees the Landesbanken had, their internationalisation increased in the first decade of the twenty-first century, including involvement in the US mortgage market. This caused major issues for several of them after the banking crisis of 2008.

In the UK and the US, local and regional banks have been going through a period of significant consolidation since the 1970s and 1980s*,† and arguably even longer in the UK, resulting in a small number of banks with extensive branch networks. That is not to say that local banks, credit unions and building societies do not exist or are on the cusp of extinction. In fact, the UK saw a growth in the popularity of credit unions in the second decade of the twenty-first century, with membership increasing from 1.66 million in 2013[50] to 2.14 million in 2019[51] with 437 credit unions.[52]

Consolidation in the number of different banks has driven a substantial reduction in the number of bank branches due to overlap of branch networks. For example, there are stories of Santander having three branches on some high streets after its acquisitive entry into the UK. Closures were inevitable. This, alongside the growth first in telephone banking and then later in online banking services, led many in the banking industry to believe that the concept of the local, branch-based retail bank was all but dead, although many also recognised that the size of a bank's customer base remained proportional to its branch footprint.

However, in 2008, Vernon Hill, following the sale of Commerce Bancorp in the United States, decided to enter the UK market. Having built Commerce Bancorp from a single branch in 1973 to 440 branches in 2007,[53] Vernon decided that his model should work equally well in the UK. He launched Metro Bank in 2010 and the concept of branch banking in the UK was, if not re-born, then certainly given a new lease of life. Metro Bank is not the only bank that recognises the branch as the core of its business. Handelsbanken, a Swedish bank with branches in the UK (and in Sweden, Denmark, Finland, Norway and the Netherlands),[54] has its entire model centred on the branch concept, making individual lending decisions at the branch level.

Despite these exceptions, branch banking does appear to be on the decline, with most banks seeking to retrench their branch networks and focus on high footfall locations. For many villages and small towns in the UK this is resulting in the closure of all branches.

2.4 Banking Consolidation

While there is a degree of smaller-scale local banking (such as credit unions and building societies and small-scale banks), in most geographies it is also true that much of the banking by value is concentrated in just a handful of banks in many countries. In the UK

*The UK had hundreds of local savings banks (termed trustee savings banks) in the nineteenth century. By 1975 there were 73, which were consolidated due to government action into 20 and then 16 regional savings banks. In 1983 these consolidated again into TSB England and Wales, TSB Northern Ireland, TSB Scotland and TSB Channel Islands. TSB England and Wales became a listed entity, TSB Group plc, in 1986, with TSB Northern Ireland and TSB Scotland fully owned subsidiaries. Allied Irish Banks bought TSB Northern Ireland in 1991 and TSB Channel Islands became a subsidiary of TSB Group in 1992. TSB Group merged with Lloyds Bank in 1995. Following UK government support during the 2008 credit crisis, the European Commission forced Lloyds to divest 632 branches which were formed into TSB Bank plc in 2013 (see reference 47). Spain's Banco de Sabadell purchased TSB Bank plc in 2015.

†There were 952 building societies in the UK in 1940, 481 in 1970, 273 in 1980, 101 in 1990, 67 in 2000, 49 in 2010 (see reference 48) and 43 in 2020 (see reference 49). Of those remaining, one building society had over 50% of the sector's total assets.

and US the consolidation of the banking industry, particularly in the last 40 years, has created banking entities of huge scale and scope. For example, we calculated from data publicly available from the Federal Deposit Insurance Corporation (FDIC) that 13 banks in the US held over 50% of all domestic deposits as of the end of June 2019. Similarly, in Great Britain in 2016 over 70% of personal current accounts were held with just four banks.[55]

2.4.1 Consolidation in UK Banking

Arguably the origins of the consolidation in British banking can be traced back to the Banking Co-Partnership Act of 1826. The act was made by the Parliament of the United Kingdom in response to the financial crisis of 1825, which had resulted in the failure of 93 banks in England and Wales (approximately 15% of the total).[56] Since 1709 until this act, banks that issued banknotes in England and Wales other than the Bank of England could not be incorporated and had to be run as unlimited liability partnerships with no more than six partners, and were called private banks. Following the act, such banks could still not be incorporated, but the number of partners allowed was unlimited for banks outside a 65-mile radius of London. The government's intent was for such *country banks* to increase the number of their partners by adding a few more partners, but in reality following the act new banks were established with hundreds of partners, which were in effect joint stock companies, i.e. companies that issued shares (to their partners) and for which the owners of the shares had unlimited liability[57] – although the individual risk was more diluted, given the much larger number of partners. By 1833, 34 joint stock banks had been successfully formed and were operating, and the Bank Charter Act of 1833 confirmed that it was legal to establish joint stock banks within the 65-mile radius of London, as long as they did not issue banknotes.*,[58]

The primary benefits of the act of 1826 were the establishment of the ability for banks to raise capital from a large group of shareholders (and use that to expand their businesses) and the ability for banks to merge with and acquire each other. Further legislation in 1858 and 1862 enabled joint stock banks to give limited liability to their shareholders. From 1826 until the early twentieth century this resulted in a shift in banking activity from private banks, which peaked at nearly 800 banks in 1813 and dropped to fewer than 100 in 1900, to joint stock banks. The number of joint stock banks peaked at 122 in 1875 and then reduced as they merged and acquired each other. At the same time, the number of bank branches grew from around 1,000 in 1813 to over 6,000 in 1900[59] and then 8,610 in 1913 with a total of 88 banks (both private and joint stock).[60] As Newton points out in her paper "Change and Continuity: The Development of Joint Stock Banking in the Early Nineteenth Century":

> *The absorption of small-scale joint stock and private banks into a national network of branches owned and controlled by large-scale, limited liability clearing banks was more or less complete by 1918.*[61]

*The Banking Co-Partnership Act of 1826 had actually left it open whether joint-stock banks that did not issue notes could be established within a 65-mile radius of London. After the 1826 act, the government obtained a ruling from the Court of King's Bench that they were legal, and the 1833 act regularised the situation.

The clearing banks were so called because they operated the cheque-clearing systems in England and Wales and in Scotland. Other banks and financial services businesses could only access cheque clearing via the clearing banks. The clearing banks also provided funding to smaller banks and other financial institutions. The situation described by Newton persisted for several decades, including through another world war, although consolidation continued. For example, 13 banks or banking groups owned the 16 clearing banks in England and Scotland in 1960.* At the start of 1970, although there were still nominally 14 English and Scottish clearing banks (according to the Bank of England),[62] these were owned by six banks or banking groups,† and during the course of 1970, the number of English and Scottish clearing banks dropped to 10 due to the consolidation of several by their owners.‡

Starting in the 1960s, though, the UK government came under growing pressure to increase competition in banking in response to complaints about the clearing banks' conservative lending policies (which in turn were heavily influenced by individual lending caps the Bank of England imposed on the clearing banks and other institutions) and their operation of a cartel.[63] Government action in the market and regulatory and legal changes seemingly abounded in the next two decades.

The UK government established the National Girobank in 1968 (initially called the Post Office Giro), which was seen as a means of providing a cheap and efficient money transfer system for retail and business customers through the existing network of post offices.[64] Although it took several years to establish itself, it became successful through providing cheap cash deposit facilities for businesses, distribution of government benefits and, from 1978, a free interest-bearing personal bank account[65] that spurred the large UK banks to respond in kind. At its peak it was the UK's sixth-biggest bank. It was

*The 16 London and Scottish clearing banks in 1960 comprised 12 banks or banking groups. Eight were independent London clearing banks (Barclays Bank, District Bank, Lloyds Bank, Martins Bank, Midland Bank, the National Bank, National Provincial Bank and Westminster Bank). There were three London clearing banks owned by other banks (Coutts & Co, owned by National Provincial Bank; Glyn, Mills & Co, owned by the Royal Bank of Scotland; and Williams Deacon's Bank, also owned by the Royal Bank of Scotland). There were four independent Scottish clearing banks (Bank of Scotland, Clydesdale Bank, National Commercial Bank of Scotland and the Royal Bank of Scotland Limited) plus The British Linen Bank which was owned by Barclays Bank but operated separately from it.

†At the start of 1970, there were 14 London and Scottish clearing banks which comprised six banks or banking groups. There were three independent London clearing banks (Barclays Bank, Lloyds Bank and Midland Bank). National Westminster Bank owned four London clearing banks (Coutts & Co, District Bank, National Provincial Bank and Westminster Bank) and the Royal Bank of Scotland owned three (Glyn, Mills & Co, the National Bank and Williams Deacon's Bank). There were two independent Scottish clearing banks (Bank of Scotland and The Royal Bank of Scotland) and two that were owned by other banks (Clydesdale Bank was owned by Midland Bank and The British Linen Bank had been sold in 1969 by Barclays Bank to Bank of Scotland for a 34.5% stake in the combined group, which Barclays Bank retained until 1985).

‡By the end of 1970 there were four independent London clearing banks (Barclays Bank, Lloyds Bank, Midland Bank and National Westminster Bank) and two that were owned by other banks (Coutts & Co was owned by National Westminster Bank and Williams & Glyn's Bank was owned by the Royal Bank of Scotland). There were still two independent Scottish clearing banks and two that were owned by other banks, although The British Linen Bank was absorbed into Bank of Scotland on 1 March 1971.