Managing the central bank's balance sheet: where monetary policy meets financial stability

In this lecture,[1] *Paul Tucker*[2] *sets out the analysis underlying the Bank's announcement on 22 July of major reforms to its operations in the sterling money markets. He outlines the problems with the current framework and goes on to argue that the new system represents a fundamental change in how the Bank thinks about the implementation of monetary policy, the lubrication of the wholesale payments system, and the provision of liquidity insurance to the banking system, all of which are facets of the management of central bank money and the central bank's balance sheet.*

When, ten years ago, Mervyn King delivered a lecture to mark the fifth anniversary of Tim Congdon's Lombard Street Research, he reviewed ideas on the monetary transmission mechanism and, engaging with part of Tim's substantial contribution over many years, the role of money (and credit) within it. These days most such accounts—including that published by the MPC[3]—begin with a simple assertion that the central bank sets the short-term nominal interest rate. And they go on to explain how, given sticky wages and prices, that enables the central bank to shift the short-term *real* interest rate in a way that either restrains or stimulates aggregate demand.[4] Notice no mention of money here. On this view of the world and, in particular, given this way of *implementing* monetary policy, money—both narrow and broad—is largely endogenous. The central bank simply supplies whatever amount of base money is demanded by the economy at the prevailing level of interest rates. Depending on the stability of the demand for money, the monetary aggregates can be useful *indicators* of what is going on in the economy,[5] but they are not necessarily *doing* anything causal. This apparent relegation of money in policy debates often troubles policymakers like me who emphasise that monetary policy's main capability is to deliver a medium-term path for *nominal* variables; and it does not, in fact, logically preclude the existence of an effect working through the quantity of money, via liquidity and other relative risk premia etc.[6] But, in any case, it should not obscure the fact that the very first step of the transmission mechanism—setting a short-term nominal rate—turns precisely on how we manage access to our money. There is a curious lack of interest in how this is done; and occasionally some puzzlement. Indeed, a former Chief Economist of the Bank, Christopher Dow, ended up concluding that it was just some miraculous convention that the banks chose out of politeness to follow.[7]

Well, last week we announced plans for the biggest shake-up in how we implement monetary policy for at least a quarter of a century. I want to use today's occasion to explain the analysis underlying these reforms. This will involve coming clean about how compromises with the first Thatcher government, during the monetary base control debate of the very early 1980s, had the unfortunate effect, albeit with a lag, of clouding the Bank's thinking about the feasible role of open market operations in the framework for setting

(1) Delivered on 28 July 2004 to mark the 15th anniversary of the founding of Lombard Street Research. The views expressed are those of the author and do not necessarily reflect those of either the Bank of England or other members of the Monetary Policy Committee. My thanks to the team that has planned the changes to the Bank's market operations, led by David Rule, Sarah Breeden and Niki Anderson of the Sterling Markets Division. My profound thanks to Roger Clews, who is truly a co-author of this paper and of many of the ideas in it; and my thanks to Kath Begley and her colleagues in the Bank's Information Centre for archival support for Roger's historical research. Special thanks also to Peter Andrews, who was the first amongst us to see that the big issue was whether to remunerate reserves. In addition to them, I am grateful for comments from the Governor, Andrew Bailey, Charles Goodhart and, also for research support, Fergal Shortall. And, finally, as ever, my thanks to Sandra Bannister for secretarial support.

(2) The Bank's Executive Director for Markets and a member of the Monetary Policy Committee.

(3) 'The transmission mechanism of monetary policy', by the Monetary Policy Committee reprinted in *Bank of England Quarterly Bulletin*, May 1999, pages 161–70.

(4) An early account is *Interest and prices*, by Knut Wicksell, London: Macmillan (1898), 1936.

(5) See, for example, 'Money and credit in an inflation-targeting regime', by Andrew Hauser and Andrew Brigden, *Bank of England Quarterly Bulletin*, Autumn 2002, pages 299–307.

(6) As discussed, for example, in *Money and the economy: issues in monetary analysis*, by Karl Brunner and Allan H Meltzer, New York: Cambridge University Press, 1993. It's just that we do not know how to identify or quantify such elements of the transmission mechanism.

(7) See *A critique of monetary policy: theory and British experience*, by J C R Dow and I D Saville, Oxford: Oxford University Press, 1988, page 217.

interest rates. On a more positive note, I shall also outline how, operationally, our monetary and financial stability roles fit together.

Managing central bank money: demand for reserves and the shape of a central bank's balance sheet

Both missions stem from the special nature of our liabilities: central bank money. We are able to implement monetary policy because the economy has a demand for central bank money and, as monopoly suppliers, we can set the terms on which we provide it. The demand for our money is manifested in two ways—holdings of notes, and bankers' balances with us. This reflects structural features of the financial system and, in particular, the way risk is managed in a fractional-reserve banking system.

Although some payments are still made using our notes, most are made in commercial bank money (through transfers of deposits). But deposit money is subject to risk. Commercial banks are in the business of providing liquidity insurance to their customers—via deposits withdrawable on demand and via committed loan facilities—and, as such, are themselves inherently susceptible to liquidity crises. In consequence, customers want to be assured that banks can maintain convertibility into central bank money (notes). And banks therefore have to manage their various credit and other risks, including 'reinsuring' against their liquidity commitments. Second-tier banks can try to acquire such reinsurance by paying for lines of credit from the largest banks.(1) But the largest banks cannot buy liquidity insurance from each other without incurring an unacceptable level of (contingent) counterparty credit risk. They have to self insure, which they do by holding high-quality assets that can be exchanged at the central bank for 'cash'—or, rather, for a credit to their account at the central bank.(2)

That relates to the second source of demand for central bank money: bankers' balances. For commercial bank money to be used as a means of payment, banks have to settle transfers of deposits amongst themselves. The big banks—ie the so-called settlement banks—settle in Bank of England money, and to that end maintain balances with us. Why is that? If they settled in each other's money, the consequent credit exposures would not be controllable—intraday or from day to day. To avoid that, they settle payments in the 'final settlement asset', central bank money. This makes the system as a whole safer. (It isn't some newfangled thing, by the way. Since the 1770s,(3) the banks have had increasingly formal arrangements to settle the clearings in Bank of England money—first in notes and then, from 1854 up to today, via deposits held with us.(4))

These sources of demand for our money rely on two preconditions: the integrity of our balance sheet and, in a fiat money system, a decent monetary policy. Without them, agents might drift to using final-settlement assets which could provide an alternative unit of account for the economy. Neither is currently a worry!

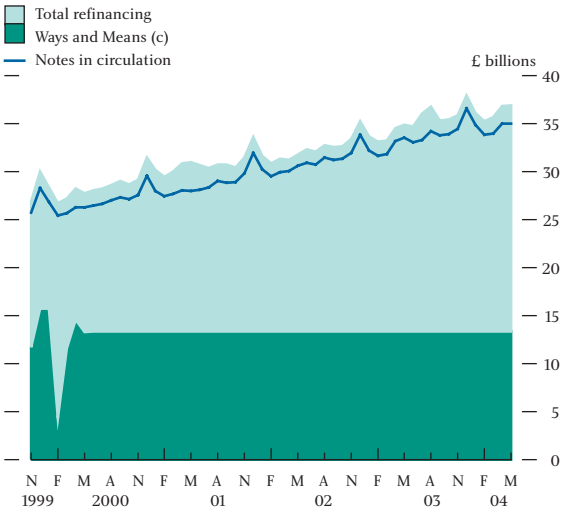
Developments in the demand for the two types of central bank money—notes and bankers' balances—drive the shape of our balance sheet: they comprise the bulk of our liabilities. For careful students of the Bank, I should perhaps make it clear at this point that in what follows—and, more important, in our analysis—Issue Department and Banking Department are treated in a completely joined-up way.(5) The separation was a central feature of the 1844 Bank Charter Act, which posited that convertibility of our notes into gold was a sufficient specification of a central bank's role. It did not have much merit then, amongst other things because it failed to recognise the importance of bankers' balances; is an aberration in today's fiat money system; and has not affected the high-level architecture of the system we are planning.

Broadly, as the economy grows, demand for our notes increases.(6) The banks have to buy the notes from the Bank, and they draw down their balances with us to do so. But that buffer is limited and, in consequence, they

(1) In the United Kingdom, this dates back to the second half of the 18th century when the 'country banks' banked with the 'London banks', and they banked with the Bank. The UK payment system remains tiered in that sense.
(2) This is the basis of the FSA's stock liquidity requirement for the largest UK banks, which was introduced in 1996. See Box 4, 'The sterling stock liquidity requirement', in 'Banking system liquidity: developments and issues', by Graeme Chaplin, Alison Emblow and Ian Michael, *Financial Stability Review*, Bank of England, December 2000, pages 93–112.
(3) And so well before Bank of England notes became legal tender (1833) or we were granted a monopoly on note issue (1844).
(4) *The Banker's Clearing House*, by P W Matthews, London: Pitman, 1921.
(5) This is also reflected in the presentation of a consolidated summary balance sheet for the Bank in the final section of the *Quarterly Bulletin's* regular 'Markets and operations' article, which we have tried over the past couple of years to make more complete and transparent.
(6) One recent study of the demand for notes is 'Assessing the stability of narrow money demand in the United Kingdom', by Kathryn Grant, Gertjan Vlieghe and Andrew Brigden, *Bank of England Quarterly Bulletin*, Summer 2004, pages 131–42.

are 'structurally short', with the Bank having to lend to them (or buy assets from them). At an aggregate level, we do that via open market operations (a term originating from around a century ago).[1] So, in line with double-entry bookkeeping, both sides of our balance sheet expand as the demand for our notes expands (Chart 1). For what are expected to be permanent increases in the value of notes in issue, the central bank can in principle provide the necessary assistance—ie inject reserves—by buying long-maturity assets, which would be done at market rates. Shorter-term assistance rolls over more frequently, and is typically used by central banks to provide reserves at a rate in line with the policy rate.

Chart 1
Bank notes in circulation and the stock of refinancing(a)(b)



(a) The difference between notes and total refinancing is accounted for by liabilities arising from bankers' balances, the Bank's capital, central bank deposits etc.
(b) Monthly averages.
(c) An illiquid advance to HM Government. This fluctuated prior to the transfer of responsibility for UK central government cash management to the UK Debt Management Office in April 2000. The Ways and Means balance is now usually constant, varying only very occasionally.

Meanwhile, the level of end-of-day balances that bankers want to maintain with the Bank is driven by (i) the rate we pay, which in the past has always been zero, and the rate we charge on overdrafts; and (ii) the precision with which they can manage their payment flows over a day as a whole. The greater their control over payment schedules, and the more effective the Bank is in ensuring that the system as a whole is square, the smaller the end-of-day buffer the banks need in the form of balances with the Bank. The fact that such balances are unremunerated has, in practice, provided a powerful impetus to end-of-day payment system efficiency. The whole system now rests on the banking system targeting *aggregate* balances of just £45 million (*m*illion not *b*illion), compared with average daily flows in the CHAPS payment system of over £150 billion (more than three thousand times greater). So although the implementation of monetary policy *does* depend on banks' demand for our money, they don't seem to demand very much of it! In the framework we are moving to, the influences on the demand for notes will be unaffected, but the bankers' balances regime will change materially.

Outline of the current and new systems

Where are we moving from, and to? First the current system. Its essence[2] is that the dozen or so sterling settlement banks have to maintain non-negative balances with us at close of business each day (the £45 million mentioned earlier). They receive no interest on positive balances, but incur a penalty rate if overdrawn. The system needs to borrow from the Bank so that the settlement banks can meet their target balances and, thereby, avoid the penalty charge on overdrafts. To that end, each day we publish a forecast of the system's shortage and undertake to make the system square (ie to achieve the £45 million target). Open market operations (OMOs) with a two-week maturity are conducted each day at the MPC repo rate; the stock of OMO loans outstanding has in recent years typically been around £20–£25 billion. There are two rounds of OMOs (9.45 am and 2.30 pm) to cater for updates during the day to our forecast of the shortage. Towards the end of the day, there are overnight lending and deposit facilities—broadly, for settlement banks/OMO counterparties—to be sure that the system is 'square'; these facilities carry penalty rates in order to induce participation in OMOs. Not all settlement banks are OMO counterparties, and *vice versa*. OMOs can span MPC dates, so we can have OMO loans outstanding which carry a different rate from that most recently decided by the MPC.

The new system will work as follows. A broad range of banks, including all of the settlement banks but going beyond that group, will agree to hold a specified positive balance with the Bank *on average* over a maintenance period lasting from one MPC meeting to the next. The level of balances targeted will be chosen by individual banks: *voluntary reserves*. For the first time in its history,

(1) *The Bank of England 1891–1944 Volume 1*, by R S Sayers, Cambridge: Cambridge University Press, 1976, page 28.
(2) For more detail, see the 'Red Book', www.bankofengland.co.uk/markets/money/stermm3.pdf.

the Bank will pay interest on such reserves: at the MPC's repo rate. The consequent demand for reserves will be met via a weekly OMO with a maturity of one week, and a fine-tuning repo on the final day of the maintenance period. There will be standing lending and deposit facilities available all day to banks generally. On the final day of the maintenance period, these facilities—used if a scheme member would otherwise under or overshoot the target—will carry rates ±25 basis points from the MPC's repo rate. Earlier in the maintenance period, the penalties on the standing facilities will be higher, perhaps ±100 basis points. If short-term OMOs were ever to span an MPC meeting, we envisage that the rate charged would be indexed to the MPC's rate.

All that sounds—and is—rather technical. But, compared with the past couple of decades, the plan reflects a fundamental change in how the Bank *thinks* about the implementation of monetary policy and the management of our balance sheet more generally, including how we support the stability of the system. This will, I hope, begin to become apparent by my explaining why we *need* to move away from where we are now.

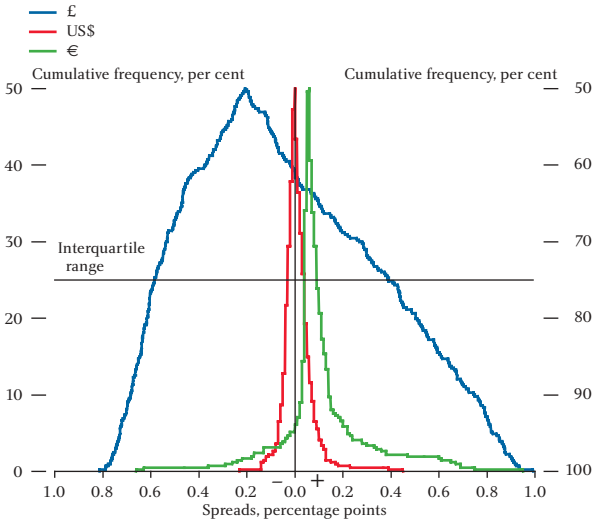
Problems with the current system: the need for reform

There are three types of problem with the current framework. First, it is overly complex: the system provides for four rounds of operations each day and on most days there are at least two. The end-of-day arrangements are especially elaborate. Second, when the MPC is expected to change rates, the ultra-short maturity rate structure 'pivots' in a rather perverse way, because the daily two-week repos span the MPC meeting but are conducted at the existing rate.(1) For all but the initiated, this makes it harder to decipher expectations from ultra short-term money market rates.

And third, the overnight rate is highly volatile by international standards—from day to day, and intraday (Charts 2 and 3). Although it has typically not affected longer-maturity money market rates and so has not impeded the monetary transmission mechanism,(2) this

Chart 2
Overnight interest rates and policy rates—United Kingdom, United States and euro area

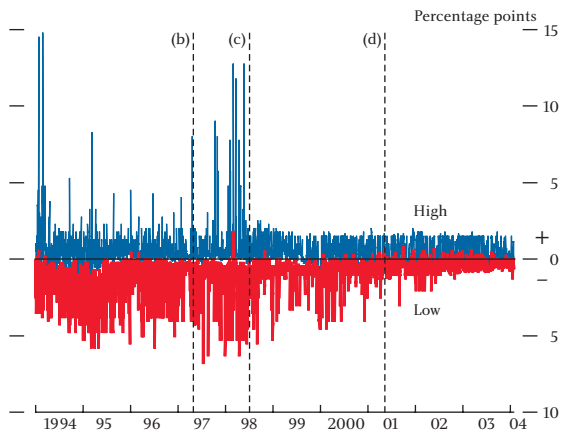
Chart 3
Cumulative folded distributions of overnight/policy rate spreads (2 January 2002 – 8 August 2003)



volatility has troubled the Bank for some years. It was considerably reduced by some major surgery in the mid to late 1990s reforms (Chart 4), which eliminated the capacity of banks to 'corner' the limited amount of eligible collateral then available (see Annex 1). But those reforms did not eliminate the capability of those large sterling banks that are OMO counterparties to move the overnight rate around, typically resulting in periods of persistent *softness* designed to reduce the cost of financing longer-maturity assets. This was a frequent occurrence until mid-2003 when the Bank's concern became evident.

(1) Arbitrage tends to make expected overnight rates over the relevant two-week period equal to the rate at which the Bank lends in its operations. So, if the Bank lends at the MPC's existing repo rate and pins down the market rate at that maturity, overnight rates up to the meeting will fall below this rate if the MPC is expected to raise its rate at its next meeting. Bidding for two-week money in the Bank's regular OMOs also rises. Conversely, if the MPC is expected to reduce the Bank's rate, overnight rates up to the meeting date rise and participation in the two-week operations falls.
(2) See, for example, 'Money market operations and volatility in UK money market rates', by Anne Vila Wetherilt, *Bank of England Quarterly Bulletin*, Winter 2002, pages 420–29.

Chart 4
Volatility of the sterling overnight interest rate(a)



(a) High and low of the day observed by the Bank's dealing desk as a spread to the policy rate.
(b) Daily gilt-repo OMOs introduced.
(c) End of transition for discount houses.
(d) Introduction of the deposit facility.

Also, since the discount houses withered away in the 1980s,(1) there has been less active market-making in overnight money. In consequence, there can be temporary frictions in the distribution of 'reserves' when it is not the large banks/OMO counterparties that are short but rather a range of smaller banks. The result has been occasional *tight* overnight market conditions, with market rates moving towards the Bank's penal 'late lending' rates.

A number of foreign banks, securities houses, corporate treasurers and money managers have told us that these characteristics of the sterling money markets deter their full participation. The volatility may, for example, have impeded the development of the overnight interest rate swap (OIS) (derivative) market relative to, say, the euro OIS market.

Less tangibly, but importantly, the reputation of the sterling money markets is impaired. How come, some ask, that the Bank has what is regarded by many outside commentators as a state-of-the-art monetary policy regime, is amongst the leading official institutions in financial stability analysis and surveillance, but has such a peculiar core money market?

So we *need* reform.

Things *were* worse before the mid to late 1990s reforms. Then the Bank had to stop relying on a moribund market (the bill market) and atrophied institutions (the discount houses); the Bank created the gilt repo market (now with nearly £200 billion outstanding), transitioned out the discount houses, and introduced a (wide) interest rate corridor for the first time. We do not have an equivalent crisis now, and we have therefore been able to step back and consider the framework more fundamentally and in the light of our high-level objectives. We made those objectives clear in the paper we published on 7 May (Annex 2).

Maturity of rates targeted

The primary objective is to stabilise short-term rates at the policy rate. Up to now, there has probably been a fairly widespread perception that, by conducting OMOs at a two-week maturity, the Bank has aimed to steer, or even set, money market rates at a two-week maturity. Prior to 1997, the Bank's daily OMOs routinely included outright purchases of bills out to a maturity of one month, and occasionally three months, encouraging a perception that the Bank wished to steer rates at those maturities. In fact, the maturity of a central bank's OMOs and the maturity of the rate(s) it aims to steer/set need not be bound together in that way. At times, a clear distinction does not seem to have been made between the maturities at which the Bank sought to *set* rates and the maturities at which the market rate would be determined by market *expectations* of the future path of the official policy rate.(2) Under monetary regimes where policy decisions were not taken and announced regularly, such a distinction was harder to make as the horizon to the next decision was uncertain.

But whatever the validity or otherwise of the idea that the Bank was, under past regimes, using its operations to steer rates at two weeks, one month or whatever, it is not what we should be doing now. And, in truth, it has been less appropriate since the early 1990s, when in a series of steps the authorities introduced the system of deciding and announcing the level of the official interest rate on a regular monthly timetable. All we can, and should, do is set the interest rate up to the next MPC meeting—a period of up to a month or so immediately

(1) The houses existed until the late 1990s, but with their presence in the market a shadow of earlier decades. They had been key intermediaries in the money markets since the mid-19th century, when what later became known as the clearing banks effectively outsourced their treasury management operations via holding secured deposits with the houses. Those deposits were run up or down as the clearers had surplus or deficit liquidity. The clearing banks progressively reclaimed their treasury function during the 1980s.
(2) As the archival research reported later in this paper demonstrates, for much of its history the Bank did make the distinction. Blurring seems to have resulted from the peculiarities of the regime introduced in 1981. See below and Annex 3.

after an MPC meeting, but eventually of just one day. Beyond the next MPC meeting, money market rates should be determined by what market participants expect the MPC to decide. Hence our objective is to have a basically flat curve, at the MPC's rate, out to the next MPC meeting. The market in overnight money would then be used by banks for liquidity management but not to speculate on the rate.

Amongst other things, this means eliminating pivoting when an MPC rate change is expected. That should be straightforward. It entails not operating *at a fixed rate* beyond the next MPC meeting, which can be achieved either by simply not operating at all beyond the next MPC meeting, or via any such operations being at a market-determined rate or indexed to the prevailing MPC rate.(1)

I have not yet quite specified which market rate we are targeting. In the sterling markets, central bank money and commercial bank money are exchangeable at par, and so one cannot identify a 'market interest rate on central bank money'. Rather, we are interested in influencing, via arbitrage, the rates on those money market instruments carrying the lowest possible credit risk, with the market determining credit risk premia on other instruments and transactions.

In routine circumstances, the Bank is not trying directly to influence the price of assets taken as collateral in our open market operations, which these days are effected via repo. Although technically a purchase and resale of securities, the securities exchanged in our repos play no role in setting policy. They are nevertheless vital, as they constitute the collateral securing the Bank's credit exposure to its counterparties.(2)

How the central bank sets rates

In terms of the overriding objective of stabilising ultra-short interest rates at the MPC's rate, the key is to ensure that we are both the marginal supplier and taker of 'reserves'. In theory, there are two possible ways of achieving this. One is to use OMOs to adjust the quantity of reserves to bring about the desired short-term interest rate, implicitly or explicitly drawing on an identified demand schedule. Neither in the past nor in the current review have we even briefly entertained the notion that this is realistic.

The alternative way for the central bank to establish itself as *the* rate-setter is to be prepared to supply (or absorb) whatever liquidity the market demands at its chosen rate(s). The most precise way of doing this is through so-called 'standing facilities' in which the central bank lends (secured) whatever is demanded at a fixed rate or takes on deposit whatever is supplied at a fixed rate.(3)

This points to the underlying problems with the Bank's current system. On their own, OMOs are *not* sufficient to make the Bank the rate-setter if, as now, they are used simply to offset the market's *net* liquidity need given a specified maintenance requirement. This is the *first fundamental flaw* of the Bank's current system. Broadly, at present a single OMO counterparty can take our money—so that the system is square *vis-à-vis* the Bank—and seek to influence the market overnight rate by trading at a different rate from the Bank's rate (up to the boundaries formed by the current *wide* corridor). A mistaken emphasis on OMOs as 'setting the rate', rather than on standing facilities, has been one precondition for the volatility in the sterling overnight rate. So a *first basic design principle* is that a well constructed system involves the *possibility* of *gross* intermediation across the central bank's balance sheet.

Divergence of the market rate from the MPC's rate can be caused by a maldistribution of liquidity among institutions (accidentally so or through deliberate hoarding). If the market rate diverges from our rate, the banks that are having to pay/receive the 'wrong' rate should be able to come to the Bank. At present, however, only relatively few banks—the settlement banks and OMO counterparties—have access to the Bank's standing facilities. This is the *second fundamental flaw* of the current system. Furthermore, the penalty on

(1) The Bank provided for indexed repos over Y2K (for a description see 'Sterling market liquidity over the Y2K period', *Bank of England Quarterly Bulletin*, November 1999, pages 325–26).

(2) Eligible collateral has to meet two tests. It should be high-quality, and to that end we recently supplemented our criteria with a public ratings cut-off of Aa3. And there should be plenty of it. Because we take EU government securities, there is some £3 trillion outstanding—somewhat larger than the £20–£25 billion stock of OMO lending in recent years!

(3) It is important to stress that this view of how the central bank's rate can most expeditiously be made effective does not entail a particular set of views about whether money is 'special' and thus about how policy rate changes are transmitted into other asset prices etc. Thus, for example, the fact that Michael Woodford advocates a narrow corridor system does not entail that practitioners who adopt such systems also share a view that nothing would be changed if money gave no special benefits (such as liquidity) to its holders. This is relevant to the range of policy options available at the 'zero bound' (briefly discussed later in this paper). For Woodford's analysis, see Chapter 2 of his *Interest and prices: foundations of a theory of monetary policy*, Princeton: Princeton University Press, 2003.

intermediating via the Bank's balance sheet should not be too great; otherwise 'victim' banks may prefer the costs of rate volatility and/or persistent tightness or softness in rates. The rates on the Bank's current 'facilities' are 200 basis points apart. To have closer control over rates using standing facilities, the Bank's interest rate 'corridor' needs to be narrower. A *second basic design principle*, therefore, is that *access* to intermediation via the Bank's balance sheet needs to be widespread and at an unprohibitive price.

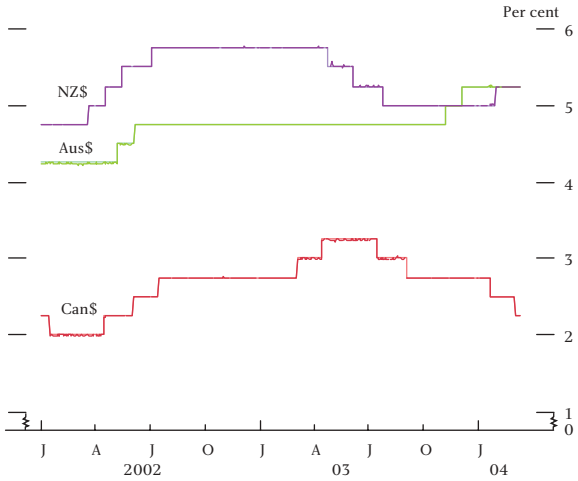
There are lots of ways of satisfying those basic design principles. At a high level of generality, the Bank could be the marginal player/price-setter in a system where banks *actually* intermediate across our balance sheet. But we could also achieve that in a so-called 'corridor system' where the rates on lending and deposit facilities provide a corridor for market rates; most intermediation occurs via the interbank market at prices *within* the corridor; and 'symmetry', which I shall explain later, delivers a market rate equal to the mid-point of the corridor, chosen of course to be the MPC's rate.

Why not have a zero corridor?

The limiting case is for the Bank to give every bank access to borrow (against collateral, o/a credit risk) or deposit in unlimited quantities overnight at the MPC's rate, ie a *zero corridor*. With identical lending and borrowing rates, there would be no (overnight) interbank market as the intermediaries could not even recover the bid/offer spread. This would distort ultra short-term money markets, and possibly collateral markets (because the Bank lends against high quality collateral and so at times would hold large amounts of it); would cause major and unpredictable day-to-day fluctuations in the size of our balance sheet; and apply no premium for the backstop liquidity insurance provided to banks via the standing lending facility.(1) Our preference is to design a framework that can achieve our monetary policy/volatility objectives while leaving open the possibility of a private market in short-term money. To achieve those goals, we do not *need* a system that entails that the Bank is the *only* intermediary in overnight money—as overseas systems demonstrate.

Other central banks have, in fact, achieved their goals for managing central bank money in a variety of ways. One group—the ECB, the Fed—use 'reserve averaging'. Indeed, it has sometimes been suggested that the fundamental flaw in the United Kingdom's current system is that it has a one-day maintenance period. Analytically, that is *not* the fundamental flaw, as I hope is clear from my earlier remarks on OMOs. And *empirically*, a number of central banks—notably Australia, Canada, New Zealand—have achieved rate stability with a same-day system (Chart 5).(2)

Chart 5
Overnight interest rates and policy rates—Australia, Canada and New Zealand



The Bank's new system will involve both. Rates will be set at the MPC's rate using the technology of a same-day narrow-corridor system. But by employing averaging, we should need a narrow corridor to steer rates only once every so often, not every day.

Averaging and smoothing the overnight rate

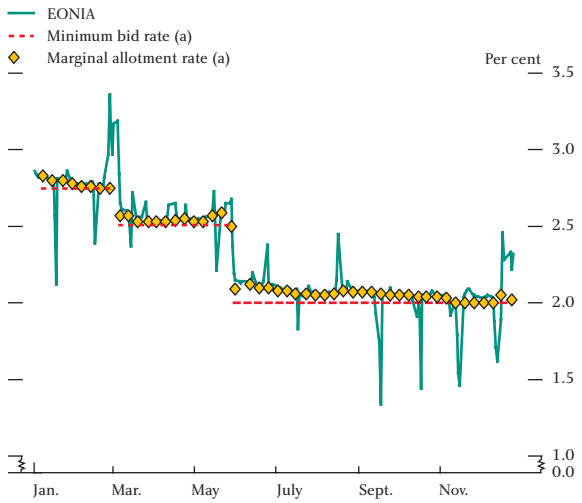
In an averaging system, a member bank has to maintain a required level of reserves on average over the maintenance period. During the maintenance period, banks are free to draw on or build up their balance at the central bank in order to meet the average required

(1) Also, the absence of a private market would potentially create scope for predatory behaviour by the clearing banks, *qua* bankers, towards other firms—securities dealers, corporate treasurers, investment firms—that bank with them, that do not have access to the Bank, and that have to manage a daily liquidity surplus or deficit. It is conceivable that there would be enough such firms for a market to exist but, unless there were active intermediaries, the search costs might be high.

(2) Comparing the US and (*inter alia*) New Zealand systems, Woodford ('Monetary policy in the information economy' in *Economic policy for the information economy*, Kansas City: Federal Reserve Bank of Kansas City, 2001) argues that same-day narrow-corridor systems are superior because they do not rely on the ability of banks to defer meeting their liquidity needs from one day to the next. In the US system, this 'deferral' capability is limited by virtue of reserve requirements being low; improvements in technology have enabled the banks gradually to reduce the level of reserves they are required to hold—a trend Woodford, like others, expects to continue. His paper does not, however, consider the possibility of attracting high levels of bankers' balances via a system of *voluntary* reserves remunerated at the official policy rate, combined with standing facilities that create a *narrow* corridor on just the final day of the maintenance period.

over the period. Such day-to-day fluctuations attract no penalty, so there is in effect no 'turn'. In principle, market rates are smoothed—a martingale is established[1]—by scheme banks varying their balances with the central bank rather than borrowing/lending in the market whenever the market rate diverges from their central expectation of the market rate that will prevail on the final day of the maintenance period; that is, the rate at which they expect to be able to 'square' up to meet their reserves target by lending or borrowing in the market on the final day. In consequence, averaging in principle establishes a flat curve through the maintenance period, with the rate expected on the final day fed back to earlier days via arbitrage. That leaves the central bank with the task of establishing its rate on the final day of the maintenance period, so that the flat curve is at the central bank's rate not some other rate. It is the same task as in a same-day maintenance system.[2] And it is also why most averaging systems exhibit UK-style volatility on their final day (Chart 6). But we believe that should be avoidable, by employing the technology of the best same-day systems.

Chart 6
Overnight interest rates in the euro area during 2003



(a) In the ECB's Main Refinancing Operation.

Setting the interest rate via a narrow corridor

Narrow-corridor systems not only put bounds on market rates, they also influence where rates will be *within* the corridor. Provided the banking system as a whole is square, the excess balances of 'long' banks by definition offset the deficit balances of 'short' banks. The former face a choice between lending in the market or depositing their excess with the central bank at a discount (say 25 basis points) to the official repo rate. The latter, 'short' banks face a choice between borrowing in the market or from the central bank at a premium (say 25 basis points) to the official repo rate. The cost of using the facilities depends on where the market rate is within the corridor. If, for example, the market rate were above the mid-point of the corridor, it would be relatively expensive to use the deposit facility but cheaper to use the borrowing facility, so banks would be more willing to run the risk of being short. They would, therefore, lend more in the market, which would tend to soften the market rate, helping to bring it back towards the middle of the corridor.

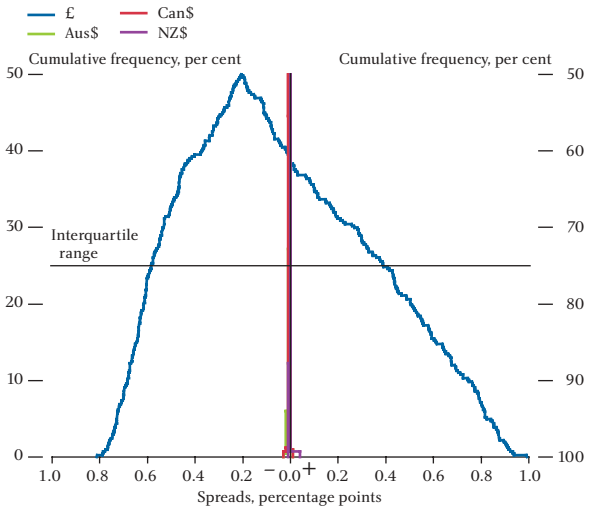
Somewhat more exactly, the pre-conditions for such symmetry are (i) central bank operations being *expected* to offset, with balanced risks, the market's net quantity shortage/surplus relative to the maintenance requirement; (ii) a market that distributes reserves efficiently; and (iii) genuine symmetry in using the two standing facilities, including no 'shame' in using the borrowing facility.[3]

In the Bank's new framework, the first of these conditions will be met via OMOs (made easier, possibly, by expressing the maintenance requirement as a small range). To help meet condition (ii) we plan to have a narrow corridor on the final day of the maintenance period. In theory, symmetry is consistent with a corridor of any width. In practice, nearly all such systems have a narrow corridor (±25 basis points). That is what we plan to employ. A narrow corridor will reduce the returns from any efforts to drive the market rate away from its midpoint. It will also reduce the cost to any potential victim banks of taking defensive action by using the Bank's standing facilities rather than the market to square their books, which would reduce the incentive for other banks to try to influence the market rate in the first place.

(1) In other words, the overnight rate on any day corresponds to the expected overnight rates on the following days of the same maintenance period.
(2) An earlier Bank analysis of averaging focused on rate smoothing and did not address how the central bank should establish *its* rate on the final day of the maintenance period. See 'Averaging in a framework of zero reserve requirements: implications for the operation of monetary policy', by Haydn Davies, *Bank of England Working Paper no. 84*, 1998.
(3) As has traditionally been thought to attach to the use of the Federal Reserve's discount window. Condition (iii) cannot be *completely* satisfied as the central bank requires collateral to cover loans. In a narrow sense, it could be made symmetric if the central bank were to provide collateral against deposits. But that opens up the possibility of such a facility being used as a general collateral stock lending facility (as occasionally happened in the United Kingdom between 2001 and mid-2003).

This kind of system has been used to great effect by New Zealand, Australia and Canada (Chart 7). Like setting policy in terms of interest rates rather than base money ('i' rather than 'm'), and inflation targeting, its basis has since been set out formally by the academic community.(1) As Keynes might perhaps have said, academic economists are often the scrupulous tidiers up after some obscure practitioner in a small country (in this case, Antipodean).

Chart 7
Cumulative folded distributions of overnight/policy rate spreads



Remunerated reserves and the scope for averaging

So the Bank of England's new system will employ both a narrow corridor and averaging. But how will we ensure that there is 'enough' averaging for the martingale to hold? The banking system's net liquidity need is affected not only by the Bank's monetary operations but also by all the other flows between us and the market, not all of which are directly controlled (notably the ebb and flow of bank notes, which the Bank supplies on demand)—these are generally referred to by central banks as 'autonomous factors'. In order to prevent scheme banks' balances at the Bank going into overdraft, another part of the Bank's balance sheet must adjust in a way that offsets movements in these autonomous factors. In principle, there are two ways of delivering this. Either the central bank conducts OMOs with sufficient frequency and size to offset autonomous flows that would otherwise put the banking system into overdraft. That is the Bank's current approach and, in effect, that of the Federal Reserve, which conducts OMOs most days. Alternatively, the central bank has to ensure that the aggregate reserves held with it are sufficient to absorb the largest foreseeable fluctuations in autonomous factors. We are adopting the second approach, which is also employed by the ECB.

For that reason we will pay the repo rate on reserves. In a system of *voluntary* reserves, anything less could lead to material fluctuations in demand depending on how our remuneration rate compared with the return on other asset classes. Since the cost of obtaining the reserves via OMOs will also be the MPC's repo rate, the demand will in theory be unlimited. We will therefore apply ceilings—possibly expressed as a percentage of so-called 'eligible liabilities',(2) and set at a level that, in aggregate, can absorb the volatility of the autonomous flows.

I should also probably note that by fully remunerating reserves, there is no tax on the banking system. There is not, therefore, the tiniest residue of schemes used elsewhere in the past designed to put a wedge between deposit and loan rates and so control monetary growth by raising the cost of bank intermediation.

OMOs and the Bank's balance sheet

It is important to be clear about what OMOs will and will not be doing in the new set-up. As now, they will *not* be used to inject a quantity of reserves according to a plan for the path of the monetary base. And they will *not* be used directly to adjust the quantity of base money to bring about the desired level of short-term interest rates. In other words, base money comprises neither a target nor an instrument of policy.

Rather, the role of OMOs will be to satisfy the system's targeted level of reserves over the maintenance period as a whole.(3) To that end, there will be a weekly repo which, reflecting feedback during the consultative period, will be for a one-week maturity. In addition, we plan to conduct a round of overnight OMOs *as a matter of routine* on the final day of the maintenance period, which will allow us to adjust for any changes in our forecast of the system's position (relative to the

(1) See, for example, Woodford, 2001, *op cit*. The New Zealand system was first described in 'Monetary policy implementation: changes to operating procedures' and 'A cash rate system for implementing monetary policy', by David Archer, Andy Brookes and Michael Reddell, *Reserve Bank Bulletin*, Vol. 62, 1999, pages 46–50 and 51–61, respectively.
(2) Broadly, a measure of the size of a bank's sterling balance sheet after netting out interbank deposits. Interestingly, the regime applying between 1971 and 1981 allowed the London clearing banks to maintain their target operational balances (1½% of eligible liabilities) *on average* over a month. This did not fit especially well with a regime in which, it seems, the Bank aimed to conduct OMOs each day.
(3) Interestingly, the ECB initially gave primacy to OMOs in describing its operations ('pivotal role in steering interest rates'), even though it would seem that that is only part of the story (*The monetary policy of the ECB*, 2001, page 65).

maintenance requirement) between the last weekly OMO and the end of the maintenance period. This *routine* fine-tuning operation is one novel feature of our plan. All short-term repos will, as now, be at the MPC's rate. That is for clarity. Technically, the rate could be determined by a tender, but we wish to rule out speculation about whether the result of a tender revealed anything about the MPC's rate intentions.

These proposed changes—and, in particular, the introduction of remunerated reserves to absorb fluctuations in the 'autonomous factors' during the maintenance period—will, therefore, affect the frequency of our OMOs, which have been daily for as long as anyone can remember. The official Bank historians report that, since the 1890s at least, the Bank placed great weight on being close to market conditions.(1) That remains as true as ever—not just of the sterling money markets but of financial markets generally, as the Governor recently underlined when discussing our Market Intelligence function.(2) To my mind, being 'in' markets frequently can aid intelligence-gathering if the activity is discretionary. By contrast, both today's OMOs, and those under the new system, are mechanical, and so conducting OMOs each day does not of itself yield intelligence. We do not, therefore, think that moving to weekly OMOs should impair the flow of intelligence to the Bank; and we shall take great efforts to stay in touch with all parts of the market so that it doesn't.

The introduction of remunerated reserves will, though, bring changes. In particular, it will almost certainly cause the Bank's balance sheet to grow. Rather than the £45 million currently held, we envisage that the banking system will hold sufficient reserves to absorb the autonomous factors—measured in terms of billions of pounds (at least).

Other things being equal, this would simply get added to the current £20–£25 billion of refinancing, provided via OMOs, which offsets the banking sector's structural short position. In fact, we will need to consider whether it will be more sensible to separate the provision of reserves needed to meet the reserves target from the offsetting of the system's longer-term short position resulting, essentially, from secular growth in the note issue. It may be that part of the latter could be injected via longer-term lending, say through the purchase of longer-term government securities. The Bank would be a rate-taker in any such official operations, which would be akin to the Federal Reserve's purchases of long-maturity Treasury bonds (sometimes known as 'coupon passes'). We will, of course, consult the market on this (if we take it forward at all). The point of mentioning it here is to make clear that, once one separates the concept of OMOs from rate-setting, it is no longer axiomatic that all official market operations should be at very short maturities. What *is* axiomatic is that the framework will be clear and transparent.

The 'classical' system; and why did we later think OMOs could set interest rates?

The role of OMOs in our plans marks a big departure from the system employed in various manifestations since 1981, which appears to have relied on OMOs to set rates. But, at least as employed, they can't. This has prompted us to examine how this state of affairs came about.

For most of its history the Bank *did* have a coherent system for setting rates, and understood perfectly well how it worked. From around the 1890s to the 1970s, the Bank employed what was generally referred to as a 'classical' system, with a somewhat penal Bank Rate which was 'made effective' from time to time by putting the market 'into the Bank'. Open market operations were conducted at market rates and were used to offset the autonomous factors—or not so used, leaving the market short and so forcing it to borrow at Bank Rate, thereby making the Bank the marginal supplier. Until 1981, the Bank did not publish a forecast of the system's shortage, and so market participants could not easily judge when, through its operations, the Bank had squared the system for the day. The Bank, which therefore had all the cards, thought of itself as controlling market rates by adjusting the scale or probability of market borrowing at Bank Rate. The rates on OMOs had *no* special significance. As Deputy Governor Harvey put it in an opening statement to the Macmillan Committee in 1930:

> '... we regard the Bank Rate [lending] as our principal weapon for carrying [that] policy into effect... open market operations... are merely part of

(1) 'The Bank had by 1890 concluded [that] the first condition for adequate influence was that the relationship between the Bank and the market should have the closeness that is consequent on frequent mutual business' (Sayers, 1976, *op cit*, page 33).
(2) In the Governor's Mansion House speech, 16 June 2004, page 350, reprinted in this *Quarterly Bulletin*, pages 349–51.

> the machinery by which the weapon of the Bank Rate is made efficient'.(1)

In 1959, Lord O'Brien and Sir Jasper Hollom—later respectively Governor and Deputy Governor but then Chief Cashier and Deputy Chief Cashier—explained the system in the same way to the Radcliffe Committee. Some of the key exchanges are set out in Annex 3 but, looking ahead to what I shall have to say about how our operational framework relates to financial stability, it is convenient to note here the sense in which Lord O'Brien referred to 'lender of last resort': 'Acting as lender of last resort, it is at Bank Rate. The other method, of buying in bills is... putting out cash in exchange for securities.'

The classical system described by Harvey and O'Brien was based on what might be called 'half a corridor'. Our reform plans are, therefore, in some respects a descendent of the classical system. We too propose to use OMOs simply to steer the quantities, and hence the probability that the market will find itself using penal facilities. But there are differences. We will not be leaving the market guessing as to whether or not we will supply sufficient liquidity to make the market square: we *will* offset the autonomous factors. And our new system will work by aiming to have the market rate in the middle of the corridor rather than, as in the classical system, by occasionally forcing the market rate to the (upper) edge of the corridor.

But the crucial point is that the *ancien régime* knew what it was doing, and didn't imagine that the OMOs *set* rates. Given the strength of induction in the Bank, where one generation learns in a critical way from another,(2) this makes the ensuing regime hard to fathom—at first sight.

Part of the answer lies in the political economy, and ideological monetarism, of the 1970s and early 1980s; and part in not adapting the framework to a profoundly altered overall monetary regime during the 1990s.

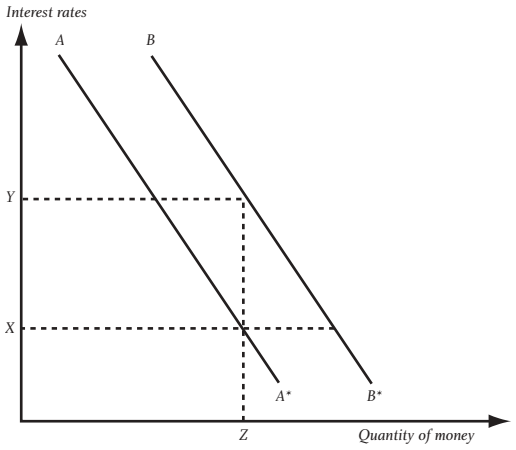
In separate acts of folly a quarter of a century or so ago, the monetary authorities sought to hide the fact that they were setting rates. In the 1970s, Minimum Lending Rate (MLR) replaced Bank Rate. This was not just relabelling, as MLR was supposed to float with market-determined Treasury bill rates, not least because that would disguise the hand of the authorities behind a tightening of credit conditions if they wanted to restrain demand. In the 1980s, a welcome emphasis on monetary variables was fallaciously argued by some to *entail* that policy should be *implemented* via a path for the monetary base, with the outcome being a messy compromise. In both episodes OMOs came to have greater apparent significance because, with Bank Rate/MLR downgraded, the authorities sometimes used the rates in a round of OMOs to institute—that is, to *signal*—a change in the market rate desired by policymakers.

The system introduced in the early 1980s after the debate on monetary base control was a particularly unfortunate aberration—rightly described by Charles Goodhart as 'confused and silly',(3) but regarded by Bank officials at the time as the best compromise they could reach given government policy that it should be consistent with transitioning to monetary base control. The resulting implementation framework was somehow meant to correct for a failure in decision-making (the 'bias to delay' in tightening monetary conditions in the face of incipient inflationary pressures). Specifically, it aimed to reduce official influence on market rates but without actually switching to monetary base control (MBC). In its OMOs, the Bank was to respond to market bids, and the so-called 'stop rate' was supposed to be no more and no less than the outcome of market clearing (although, in fact, there was an undisclosed range of acceptable stop rates agreed with the Chancellor of the Exchequer). Even when, from November 1982, it was made clear that the authorities were, after all, deciding the rate, the mechanics were left largely unchanged. In principle, the Bank was still responding to market bids, setting a rate (by lending at MLR) only intermittently.

Because the logic of the actual system was so obscure, when economists wanted to conceptualise what was going on, they used simple textbook-style models. This is nicely illustrated by Chart 8,(4) which seeks to make a

(1) *Minutes of evidence taken before the committee on finance and industry, Vol. 2*, London: HM Stationery Office, 1931, page 173.
(2) Lord O'Brien joined the Bank in 1927 and so served under both Montagu Norman and Harvey. The Executive Director under whom the current Markets area senior management first learned about the Bank's official operations was Tony Coleby, who worked in the money markets area during O'Brien and Hollom's Governorship and Deputy Governorship. Coleby's 1982 paper on the Bank's operations (see Annex 3) made clear that *'until recently...*, the operational technique for giving effect to official interest rate objectives has stayed close to the classical model' (my emphasis).
(3) 'The Bank of England over the last 35 years', by Charles Goodhart in *Bankhistorisches Archiv*, Beih. 43, *Welche Aufgaben muß eine Zentralbank wahrnehmen: historische Erfahrungen und europäische Perspektiven*, 2004, page 49.
(4) From 'The operational role of the Bank of England', by Charles Goodhart, *Economic Review*, Vol. 2, May 1985, pages 23–27.

Chart 8
The banks' demand for cash



straightforward point about the choice between controlling the price or quantity of our money. When demand for money shifts from AA* to BB*, either the extra demand is accommodated at an unchanged interest rate (*X*) or, alternatively, the price (interest rate) must rise to *Y* if supply is held fixed at *Z*. Since the Bank knew it was not operating MBC (fixing the quantity of money at *Z*), we had to be in the rate-setting version of the model, and so we *had to be* the marginal supplier of 'cash' at our chosen rate. There are many problems with this story. First, the Bank was, in fact, targeting a fixed (but adjustable) level of reserves (bank balances with us)—so we were, somehow, controlling both the rate and the quantity! Second, even when notes as well as bankers' balances are taken into account, we know that we can change the policy rate without having directly and immediately to alter the quantity of central bank money. Related to that, a further problem with the set-up is that the demand for central bank money depends not simply on the absolute level of the short-term risk-free interest rate but on where it is relative to the expected returns on other assets (the opportunity cost), which themselves may not be independent of monetary policy.

The Bank's thinking had, moreover, drifted into blurring the distinction between OMOs and standing facilities, except that the Bank thought of OMOs as modern and market-friendly. And, most important, because—as economists—we knew we *ought to* be supplying marginal liquidity at our chosen rate, we slipped into thinking that that *was*—surely had to be—what we were actually doing. So, *ipso facto*, the OMOs were setting rates.

Against that background, it is interesting that the original operational plans for the 1996–98 reforms (described in Annex 1) did, in fact, retain the insights of the classical model, with a recognition that there might be conditions in which the Bank would need to leave some of the system shortage unrelieved by the daily OMOs, forcing the market into the late lending window, in order 'to ensure that the Bank is—and is known to be—the marginal supplier of liquidity..., preventing banks seeking to substitute themselves'.(1) In the event, this discretion has not been exercised, in order to avoid the risk of any such actions being perceived, mistakenly, to convey signals about monetary policy. Quite separately from debates about OMOs, the United Kingdom moved to a monetary regime where signals via the Bank's operations were not needed and, indeed, would be counterproductive. Our reform plans have been developed with that in mind, and so with the aim that the framework itself should stabilise the market rate in the middle of the corridor rather than relying on the Bank's ability to take the rate to the corridor's edges.

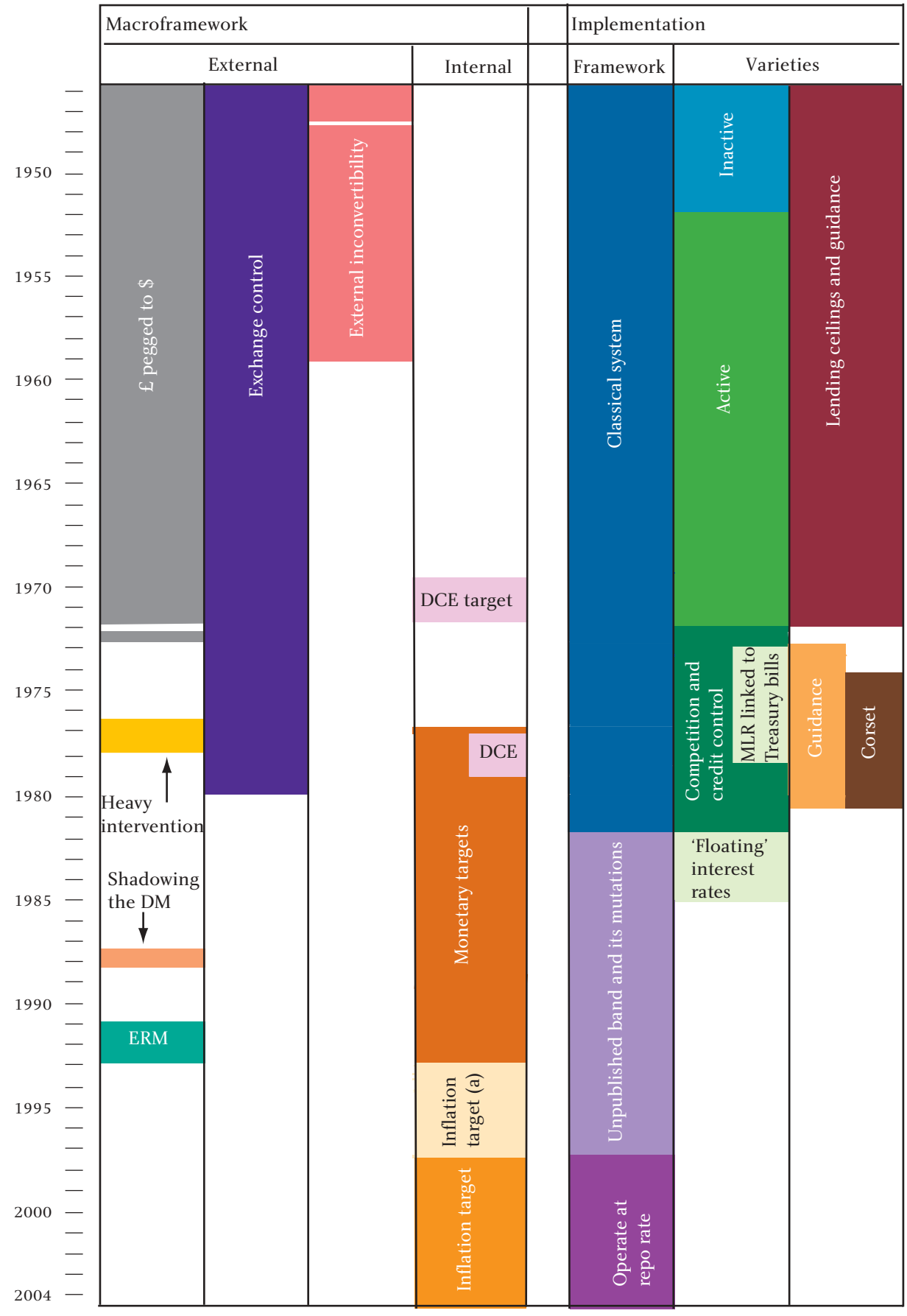
Monetary regimes and implementation frameworks

Discussion of the classical system, introduced when the United Kingdom was on the gold standard; of the confusion sown by the debate about monetary base control; and of the redundancy of using operations to signal policy in a world where the policy rate is periodically decided and announced and where policymakers' view of the monetary transmission mechanism and their reaction function are transparently communicated—all of this might imply that there has been a clear and robust relationship between the authorities' overall monetary regime and the framework for implementing policy. In fact, that does not appear to be the case, judging from Chart 9. Much the same has been found by others.(2)

This is puzzling. The classical system—of *daily* OMOs, with the options of forcing market rates up to Bank Rate and of changing Bank Rate between the Court's weekly meetings(3)—does seem reasonably well suited to the gold standard regime, with its threat of external drain. The market rate sometimes needed to be adjusted at short notice 'with the object either of preventing gold

(1) See Annex 3.
(2) See, for example, 'Instruments, procedures and strategies of monetary policy: an assessment of possible relationships for 21 OECD countries', by J Swank and L van Velden in *Implementation and tactics of monetary policy*, BIS Conference Papers, Vol. 3, 1997, pages 1–12.
(3) Known as 'a Governor's rise', which would be confirmed at the next meeting of Court (see Sayers, *op cit*, page 28).

Chart 9
Post-war monetary regimes and implementation frameworks



(a) With monetary monitoring ranges.

from leaving the country, or of attracting gold to the country'.[1] Consistent with this, Governor Norman and his colleagues are recorded as having pored over the gold position every day.

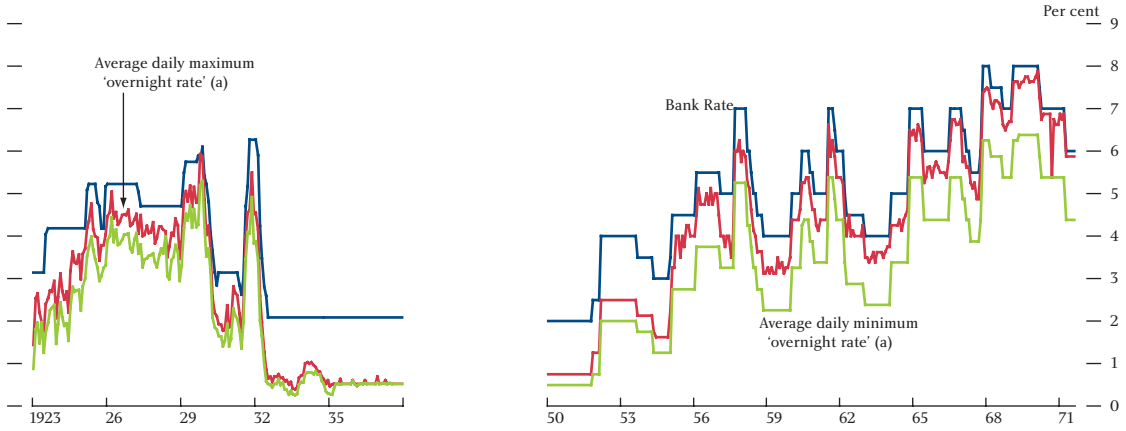
In theory at least, the same goes for any external money anchor, such as the ERM. It is striking that when the United Kingdom joined the ERM in 1990, a good deal was made of its not implying a significant change in the way monetary policy was set or implemented. It could be argued that the overnight rate needed to move to whatever level was needed to stay in the ERM band, and a daily maintenance system should, in principle, have facilitated that. In fact, that was not how policy was operated, reflecting the UK authorities' stress on the ultimate objectives of policy rather than the intermediate means of pursuing them provided by the ERM. By contrast, various other ERM members did base their policy framework on being able to influence their exchange rate via very close control of ultra short-term market interest rates—but some of them had averaging schemes, which on the face of it might have afforded them relatively little day-to-day purchase on the overnight rate.[2]

The overall historical picture is not especially coherent. I suggest that the question of whether, desirably or even optimally, there might be some mapping from monetary regimes to operating frameworks warrants research by the academic community.

In one respect, however, the evolution of the Bank's operating system *does* seem to have tracked the evolution of the overall monetary regime—the precision with which rates are set. This is another area where more research would be useful but some preliminary propositions can be advanced, if only to be knocked down. My impression from Chart 10 is that during the 1920s the Bank was fairly relaxed about the spread between Bank Rate and the market rate; that during the 1930s and the early 1950s,[3] when monetary policy was assigned a minimal role in macroeconomic management, the spread could at times be measured in terms of percentage points; and through the 1960s, when direct credit controls were employed, the authorities seem also to have been fairly indifferent to the range between maximum and minimum market rates. I have already touched on the peculiarities of the 1970s and the early 1980s. During the 1990s, and more recently, we have become more concerned about the relationship between official and market rates. I conjecture, but cannot yet demonstrate, that these patterns have something to do with the monetary regime of the day.

What *is* clear is that in our current monetary regime, we need precision. To date, the smallest change in the MPC's rate has been 25 basis points and members of the Committee each spend a great deal of time deciding the level of rates we individually want rounded to the nearest 25 basis points. The Bank's operating framework needs to deliver that precision. The new framework will make

Chart 10
Bank Rate and overnight interest rates



(a) For the pre-war period, the average daily maximum and minimum 'overnight rates' are monthly averages of the highest and lowest daily rates of interest charged on day-to-day loans in London. For the post-war period, they reflect the range of rates charged by London clearing banks for loans to the discount market on the last Friday of every month.

(1) *Interviews on the banking and currency systems*, Senate document 405, Washington: National Monetary Commission, 1910, page 26. The NMC was the body that recommended the establishment of the Federal Reserve System.
(2) In some cases, eg France, the averaging cushion had been reduced, enabling the central bank to have greater influence over ultra short-maturity rates via OMOs.
(3) There is a gap in the data during World War II.

that apparent through the overnight rate. But, as a matter of routine, we revisit the policy rate once a month, not every day. Reserve averaging seems well-suited to such a regime.

So perhaps at long last, we are planning an operating framework that matches the overall regime.

Velocity shocks and standing facilities

This brings me to the connections between the framework for implementing monetary policy, and financial stability. Here too, our planned changes should bring some significant improvements.

A key financial stability concern is to ensure that the central bank can meet increases in demand for reserves that are either system-wide or, depending on the cause, from individual banks. Either may be needed to avoid a banking system panic having systemic effects: as part of our responsibilities for providing the economy's final settlement asset, we need to be prepared to expand our balance sheet when commercial banks might otherwise be under pressure to contract theirs.(1) Neither need interfere with monetary policy. Indeed, it is important that a system-wide increase in the demand for reserves should be accommodated in order to keep interest rates stable (it is akin to a velocity shock).(2) And idiosyncratic increases in demand can be offset through adjustments to the (net) provision or withdrawal of reserves to/from the rest of the system via routine operations. A well designed framework will cater for this. The current UK system does not do so as effectively as it might.

First, it makes no provision for the banking system as a whole *routinely* to change the level of reserves that it wants to hold—except via banks acquiring more Bank of England notes from us to hold in their tills. Technically, we could increase the targeted level of end-of-day balances; and we can, of course, always simply inject reserves—'excess' to the maintenance requirement—by buying securities. But there is no routine mechanism for the settlement banks themselves to seek such an increase and they may well be deterred from doing so by the lack of remuneration. The new framework will improve on this in a number of ways. Individual banks—and so the banking system in aggregate—will be able periodically to adjust the level of reserves they are targeting. And, during a maintenance period, they will if necessary be able to hold excess reserves.

In terms of *idiosyncratic* demand-for-money shocks, we currently have specific machinery for routine liquidity provision (against high-quality collateral) only to the settlement banks (and, for part of the day, OMO counterparties); no other banks have a *routinely* available mechanism to borrow against collateral from the Bank (even at a penal rate).(3) Furthermore, the overnight lending facility is available only up to the Bank's forecast of the system's *residual net* shortage (relative to the daily maintenance requirement) after the OMOs. As well as these design faults impeding the achievement of a stable overnight rate in the ways described earlier, they may at the margin hinder a smooth response to stressed conditions. Other central banks have penal collateralised Lombard facilities (or 'discount windows') in place for *all* banks *all* of the time. Our new framework will have this feature. But period averaging permits a useful refinement. On the final day of the maintenance period, the corridor will be ±25 basis points around the repo rate, but earlier in the maintenance period the rate charged can be more penal.

(1) It turns out that this was anticipated in the early 19th century by Thornton. In a crucial passage, he argued that '... to allow of some special, though temporary, increase [in the total amount of paper issued] in the event of any extraordinary alarm or difficulty... this seems [*inter alia*] to be the true policy of the directors of an institution circumstanced like that of the Bank of England'. See *An enquiry into the nature and effects of the paper credit of Great Britain*, by Henry Thornton, 1802, edited with and introduction by F A von Hayek, New York: Rinehart, 1937. Similar points were emphasised by the Banking School 40 years later; see *Studies in the theory of international trade*, by Jacob Viner, New York: Harper, 1937 for an overview.

(2) Meeting an extraordinary large and sudden demand for liquidity may be needed to prevent the overnight rate diverging from the policy rate, for example going to the top of the interest rate corridor. It certainly does not, of itself, entail that the overnight rate *must* go to zero (or close to zero), as happened during the Federal Reserve's response to 9/11. In crisis conditions, however, the central bank may know that there is a big increase in the demand for reserves without knowing how big, and accordingly err on the side of overprovision rather than underprovision. If, *ex post*, there is overprovision, the overnight rate will tend to fall to the bottom of the corridor formed by the rate on any standing deposit facility—or to zero if there is no such facility, as in the Federal Reserve's system for example. In theory a central bank with both borrowing and deposit facilities could narrow its corridor for the overnight rate in such circumstances, in order to preserve an overnight rate in line with its policy rate.

(3) In fact, as recorded by Tony Coleby in a 1982 address to mark the centenary of the Bills of Exchange Act ('Bills of exchange: current issues in a historical perspective', *Bank of England Quarterly Bulletin*, December 1982, pages 514–18), in the 19th century: 'The Bank for its part came to understand that access to its lending facilities was a valuable privilege. Consequently, it had the problem of how to retain these facilities (and thus underpin confidence in the financial system) without giving the privileged institutions the ability to on-lend more easily and cheaply than those without access to them. The Bank's conclusion was to concentrate its lending facilities on the discount houses because they did not compete with the banks for overdrafts or other lending business....' In the planned new framework, the opposite solution will be adopted—widening access to the standing facilities to banks generally.

Partly because the way in which banks manage their liquidity affects the payment system and the routine demand for our reserves, and partly because the Bank might be called upon to supply emergency liquidity support (outside of our standard operational framework),(1) we have a legitimate interest in the way that liquidity is managed by banks and across the system as a whole. Other things being equal, our aim is to have a framework that encourages disciplined private sector liquidity monitoring and management, as a contribution to reducing the likelihood of central bank emergency intervention. That points to not giving the banks such easy/cheap access to central bank liquidity as to cause their own front-office/treasury liquidity management capabilities to atrophy or market disciplines to be eroded. And that, in turn, points to 'discount window' lending being available only at penal rates, perhaps ±100 basis points relative to repo (and to overdrafts being more penal still, as a bank in overdraft has failed to manage its liquidity during the day).

A final point needs to be added about standing facilities and the Bank's financial stability mandate. For at least 30 years, and probably longer, academic economists and other commentators have debated whether routine LOLR lending (to clearly solvent banks) should be effected via OMOs rather than via the discount window (or standing facilities), arguing that such lending should be made available only to clearly sound banks and that bilateral lending is therefore not needed: OMOs are sufficient.(2) The part of the argument about lending only to sound banks is irrelevant here as our routine operations are with counterparties that can pledge high-quality collateral. But the argument that only OMOs are needed to meet the liquidity needs of *manifestly* sound banks is flawed. In the first place, in stressed conditions with widespread nervousness about counterparty risk, a *sound* bank may create unwarranted apprehension about its position if, because of (actual or perceived) problems elsewhere, it is unusually short of funds and attempts to borrow unusually large amounts in the market, even against collateral. It can avoid risking that unnecessarily adverse reaction by using the central bank's standing facility, paying the penalty rate. The central bank does not need an informational advantage; and, because it is not itself vulnerable to a run, it is not induced to overreact as a consequence of needing to preserve reserves to maintain the integrity of its own balance sheet. Separately, and I believe decisively, the argument makes the rather splendid assumption that the money markets are always open and functioning properly. The tragic events of 9/11 underline that that cannot be guaranteed.(3) OMOs rely on markets to distribute liquidity to where it is needed. Lending facilities can do the job directly—quite apart from being needed to provide a corridor for rate setting.(4)

In short, the Bank's new system, with better-designed standing facilities, will better support *both* our monetary and financial stability missions.

In doing so, it will take us further in the direction of Bagehot's precept(5) that, so far as possible, central banks should make clear in advance their preparedness to advance liquidity, against collateral and at a penal rate, in stressed conditions. Since Bagehot's day, a lot of the central bank lending that was then discretionary has become 'hard coded' into the operating framework. As was clear from the remark of Chief Cashier O'Brien I quoted earlier, that was true of the old 'late lending' window for the discount houses; and it is true, today, of the lending facilities for the settlement banks. The new system's standing facilities will improve on those arrangements in the way I have described.

From time to time, there will be transient effects on the size of our balance sheet. In the current set-up, the *net* provision of reserves pretty well always equals the *gross* provision, because the scale of our operations is typically limited by our forecast of the system's net shortage. But that will not be axiomatic in the new world, where *gross* provision could diverge from *net* provision by virtue of the standing facilities being used. We see no difficulty with that.

(1) See the Memorandum of Understanding between HM Treasury, the Bank of England and the Financial Services Authority, available at www.bankofengland.co.uk/legislation/mou.pdf. Discretionary support operations are discussed in 'The pursuit of financial stability', LSE lecture by Governor George, *Bank of England Quarterly Bulletin*, February 1994, pages 60–66.

(2) See, for example, 'Financial deregulation, monetary policy and central banking', by Marvin Goodfriend and Robert G King, *Federal Reserve Bank of Richmond Economic Review*, Vol. 74, 1988, pages 3–22; and 'The misuse of the Fed's discount window', by Anna Schwartz, *Federal Reserve Bank of St Louis Review*, September/October 1992, pages 58–69.

(3) See, for example, 'Liquidity effects of the events of September 11, 2001', by James J McAndrews and Simon M Potter, *Federal Reserve Bank of New York Economic Policy Review*, Vol. 8, November 2002, pages 59–79.

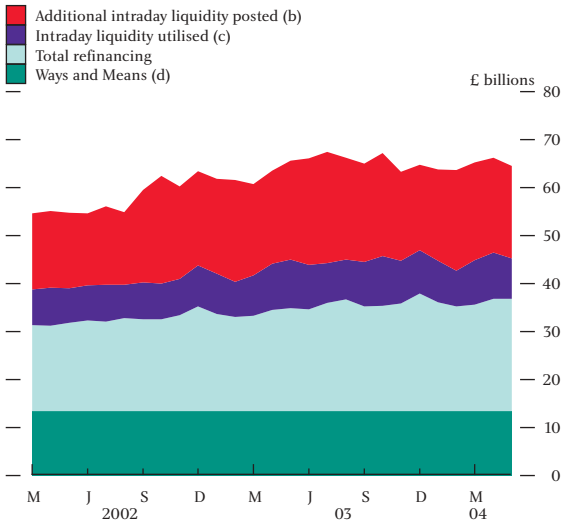
(4) For the stability of the system, it is vital that we should be able to manage liquidity even in circumstances where, for example, a business continuity problem or disaster has knocked out the European time-zone settlement systems. We therefore recently announced that, in exceptional circumstances affecting the infrastructure or firms, we will be prepared to take US Treasury bonds as collateral.

(5) *Lombard Street: a description of the money market*, by Walter Bagehot, New York: Scribner Armstrong, 1873, page 71.

The Bank's market operations, the wholesale payments system, and financial stability

Both those points—about 'hard coding' lending facilities and fluctuations in our balance sheet—already apply *during* the day by virtue of our provision of intraday liquidity to the settlement bank members of the real-time gross settlement (RTGS) wholesale payment system. All of our RTGS lending is bilateral, and so gross. As a result, our balance sheet is bigger during the day than at the close of business (Chart 11). This is all in the cause of avoiding massive daylight credit exposures amongst the settlement banks as part of their payments business, while absolutely minimising the Bank's credit risk (by taking high-quality collateral, via intraday repos).

Chart 11
Intraday liquidity and the stock of refinancing(a)



(a) Monthly averages.
(b) Monthly average of maximum intraday liquidity posted in CHAPS sterling less maximum intraday liquidity utilised.
(c) Monthly average of each day's maximum intraday liquidity utilised in CHAPS sterling.
(d) An illiquid advance to HM Government. This fluctuated prior to the transfer of responsibility for UK central government cash management to the UK Debt Management Office in April 2000. The Ways and Means balance is now usually constant, varying only very occasionally.

And, therefore, analogously with overnight standing facilities, the Bank's monetary and financial stability missions also meet in how the RTGS and 'OMO' machines fit together. The 'OMO' machine has already been described: it provides liquidity at overnight or longer maturities as part of the framework for setting interest rates. The RTGS machine determines how wholesale payments (CHAPS payments) are effected amongst the dozen or so settlement banks. All such transfers are made in real time across the Bank's balance sheet. Where a settlement bank's balance is too small to fund a payment, it borrows from us *intraday*—at a zero interest rate and in amounts limited only by how much eligible collateral it has available. In this way, the wholesale payment system is lubricated. There is squaring off at close of business every day. That is to say, if a settlement bank cannot repay its free intraday borrowing, it is charged a penalty rate on its overdraft, which is the 'bridge' to—and forms part of, now and in the future—the monetary policy machine. In terms of operational mechanisms, the two machines are joined up (settlement banks, collateral etc). In terms of pricing, the intraday and overnight 'markets' are segmented. Up to now the Bank's thinking has been that, so long as there is not a market in intraday money, we do not need to set the price in order to ensure consistency with monetary policy; and that not rationing the supply of reserves intraday promotes payment system efficiency. These are issues that we keep under review,(1) but our current reform plans maintain the segmentation.

In the new framework, liquidity will, however, be provided in a way that should help to meet the needs of the payments system as well as ensuring that monetary policy is implemented cleanly—essentially by facilitating positive reserve balances at the Bank. As in other countries' systems, it will become possible for the settlement banks to use their (remunerated) cash reserves to make CHAPS-RTGS payments during the day; ie they will be able to fund inter-settlement bank payment transfers by drawing during the day on the pot of cash they hold with the Bank, as well as by drawing on a pot of collateral to borrow from us intraday. As I've described, the cash will be provided to the *system* via OMOs; in *aggregate*, the banks will still need to borrow reserves from the Bank, but the maturity will be extended from intraday to that of the OMOs (one week). By contrast, *individual* RTGS members will have the *choice* of acquiring their reserves, and so a means of payment, indirectly via the unsecured money market rather than directly via OMOs or intraday repos with the Bank. We hope that, taken together, these measures will encourage more banks to join RTGS. In most other industrialised countries, almost *all* banks are members of the RTGS system. That eliminates intraday payments-related credit exposures, which regrettably still exist in the United Kingdom between the top tier settlement banks and their 'correspondent' bank customers and amongst the

(1) The use of our balance sheet for settlement of large interbank transfers during the day provides a bedrock of demand for central bank money.

latter. In terms of financial stability, this is a weakness in the United Kingdom's financial architecture, especially where the exposures are, or might in stressed conditions be, large.(1)

Liquidity traps and quantitative easing

I suspect that some in the audience will have been bursting with frustration at two of the assumptions that I have maintained throughout this account: that money can be regarded as largely endogenous, and that we need a system that caters only for setting interest rates. But what if interest rates were to hit the zero nominal bound, ie we were in a liquidity trap? At a policy level, the recipe would of course depend on the diagnosis of the problem; *and* on views about the transmission mechanism and, in particular, the role of money within it.(2) We would, for example, need to form a view—or individual views—on whether the injection of base money was expected to work through affecting the risk-free rate, or risk and liquidity premia on other financial assets, or both. But that lies beyond my scope today. Rather, I want to make two points about the implementation framework. First, as I have described, the Bank's new framework will make provision for the acquisition of longer-term assets (eg government securities) as part and parcel of managing our overall balance sheet in a sensible way, while making sure that the banking system is square. Faced with a liquidity trap, the Bank could in principle make purchases of securities to inject base money, which would be within our vires and technically would be 'excess reserves'. We would need to do so in a way that preserved the integrity of our balance sheet.

The second point is that, in such circumstances, there would potentially be a need to co-ordinate with government debt management, since that by definition also involves the exchange of securities for cash. In a world where OMOs were conceived as an instrument to control the path of reserves, the question of such co-ordination was familiar—as is apparent from Milton Friedman's writings.(3) Most of the time these days, it is not a practical issue.(4) But it is an issue that we have identified,(5) and there is nothing in our planned new framework that would inhibit such questions being explored if they ever needed to be.

Conclusions

In designing a new framework for the Bank's official sterling market operations, we have tried to factor in the interactions between the implementation of monetary policy, the lubrication of the wholesale payments system, and the provision of liquidity insurance to the banking system. All are facets of the management of central bank money and the central bank's balance sheet. In summary, we can implement monetary policy because we are a central bank. We are a central bank essentially because we are the bankers' bank. What we have to offer is central bank money. We have tried to take a comprehensive view of how to deploy it.

One consequence is that, in various ways, the new regime will alter the operational relationship that the Bank has with the banking system. Our aim is that lots of banks should sign up for the standing facilities, which will require little more than fixing legal agreements etc. We hope that plenty of banks—ie beyond the settlement banks—will choose to join the reserve-averaging scheme. Progress on both those fronts will be important to achieve our primary rate-setting objective. We also hope that, having opened reserve accounts with the Bank, more of the large banks will choose to become RTGS settlement banks, which would contribute to the overall safety and soundness of the financial system.

But our pre-eminent aim is to stabilise the overnight rate at the MPC's rate, so that the implementation of monetary policy is cleaner and more transparent. Working with the banking industry in the coming period of detailed planning and implementation, we are hopeful of achieving that, and so of having an implementation framework that measures up to the United Kingdom's overall monetary regime.

(1) The IMF recognised this point in its *Financial system stability assessment on the United Kingdom* (Washington: IMF, February 2003): '... the two-tier structure of the payment system may still result in significant intraday exposures between direct and indirect settling banks.... The UK authorities were encouraged to continue to give very high priority to the identification and overall monitoring of these risks', paragraph 76.
(2) For a summary of these issues see 'Monetary policy and the zero bound to nominal interest rates', by Tony Yates, *Bank of England Quarterly Bulletin*, Spring 2003, pages 27–37.
(3) See especially Chapter 3, 'Debt management and banking reform', of *A program for monetary stability*, New York: Fordham University Press, 1959.
(4) *Government debt structure and monetary conditions*, edited by K Alec Chrystal, Bank of England, 1999.
(5) The Governor discussed some of these issues in his Ely lecture, *The institutions of monetary policy*, at the American Economic Association annual meeting on 4 January 2004, reprinted in this *Quarterly Bulletin*, pages 332–45.

Annex 1

The post-1996 reforms of the Bank of England's official sterling operations

Collateral and counterparties

In the early to mid-1990s, there were two major problems. First, the range of securities eligible in the Bank's operations was narrow and values outstanding had become very small; at times the Bank held a lot of the eligible stock outstanding and the rest could, as a result, be concentrated in a few hands. Second, OMOs were conducted with the discount houses, whose capital had not grown in line with the size of the money market or, post ERM exit when the Bank had made substantial purchases of sterling, the size of the banking system's aggregate collateralised short-term borrowing from the Bank. Specifically, to control the Bank's exposure to credit risk, there were capital-related limits (referred to as Tranches 1 and 2) on so-called 'late lending' to the houses, and these limits could be lifted only by a Director (on delegated authority from the Governor). Unless the working-level operators applied for the limits to be lifted, occasionally they could constrain the supply of reserves to an amount smaller than the system's residual shortage after OMOs had been conducted (so the clearing rate was infinity!). Result: a few big clearers would, and did, shift the market rate in overnight money around at will—by holding much of the eligible collateral not already in Bank hands and by not participating in the OMOs, so that the market could, when they chose, remain very short. Occasionally the overnight rate went very high. The incredible 'supply less than demand' problem was (easily) solved in 1994–95 by allowing the Principal of Discount Office (to use the historic title) routinely to use Tranches 3–4 etc for lending to the houses. In the major reforms of 1996–98, the Bank (i) enlarged eligible collateral to include repo of gilts (and subsequently, in 1999, a much wider range of EU government securities); (ii) moved to dealing with banks and securities dealers in OMOs; and (iii) put bounds on rates via an adapted lending facility (1998) and a new deposit facility (2001).

Although analytically simple and not addressing more fundamental questions about the Bank's operational architecture, this was a major enterprise. First, the Bank had to 'create' a gilt repo market, by getting removed a whole battery of tax and regulatory impediments and by facilitating an industry code to guard against scandals of the types that had accompanied the launch of the US repo market a few years before. (The size of the gilt repo market is now approaching £200 billion.)(1) Second, the Bank had to oversee the orderly demise of the discount houses and stock exchange money brokers. Third, changes to the infrastructure were needed.

The reforms killed the ability of big banks to dominate the market's holdings of eligible collateral; and, through the deposit and lending facilities, capped volatility (see Chart 4 in the main text). But they were not enough.

Simplifications

The 1996–98 reforms also removed layers of complexity. The OMOs were thereafter conducted at the official policy rate, and the myth that the Bank was acquiescing in a 'market rate' was binned. Operations beyond the maturity of the core 'two-week' repo ceased, so that for example outright purchases of bills were confined to bills with a maturity no longer than that of the day's repo; that reduced, but did not eliminate, pivoting. The mechanism of '2.30 pm lending' at Minimum Lending Rate was scrapped. As was a facility for the settlement banks to invite the Bank to buy Treasury bills up to 3 pm each day, which was replaced by a simpler secured borrowing facility. We got rid of one round of OMOs. We increased transparency by announcing the amounts allotted in rounds of OMOs, and also the details of 'late lending'. The 'tranche system' rationing access to the discount window was abolished. And gilts of all maturities were taken as collateral rather than only gilts of up to five years. But, notwithstanding these steps, the system remained complicated.

'Clausing' and the real bills doctrine

Finally, in 2000 we abolished the requirements on the 'clausing' of bills of exchange, which had required that the underlying transaction be 'self liquidating' etc. This was a leftover from a period when the Bank genuflected in the direction of the real bills doctrine (although that seems not to have loomed as large in Bank thinking during the 20th century as it did at the Federal Reserve).(2) We got rid of clausing because we did not think it provided credit enhancement, since the Bank's claim, if the acceptor of a bill failed, would be as a general creditor of the drawer, with no lien on the cash flows from the underlying transaction.(3)

(1) See Chart 30 from 'Markets and operations', *Bank of England Quarterly Bulletin*, Spring 2004, pages 5–20.
(2) For the Federal Reserve's framework, see for example *A history of the Federal Reserve, Vol. 1: 1913–1951*, by Allan H Meltzer, Chicago: University of Chicago Press, 2003.
(3) We have recently learned that there had been an attempt to abolish 'clausing' for similar reasons in 1971 by Andrew Crockett (then a junior official in Discount Office, later a senior IMF official, Executive Director of the Bank, and General Manager of the BIS).

Annex 2

Objectives of the Bank of England's operational framework

- *Objective 1*: Overnight market interest rates to be in line with the MPC's repo rate, so that there is a flat money market yield curve, consistent with the official policy rate, out to the next MPC decision date, with very limited day-to-day or intraday volatility in market interest rates at maturities out to that horizon.

- *Objective 2*: An efficient, safe and flexible framework for banking system liquidity management—both in competitive money markets and, where appropriate, using central bank money—in routine and stressed or otherwise extraordinary conditions.

- *Objective 3*: A simple, straightforward and transparent operational framework.

- *Objective 4*: Competitive and fair sterling money markets.

Annex 3

Some historical insights into the Bank of England's official sterling operations

This annex sets out some of the source materials for the historical parts of the main paper.(1)

In his biography of Governor Norman, Henry Clay records that:

> 'In the last resort the Governor could control the supply of money in the Money Market by varying Bank Rate—its price in last resort—and by influencing its amount by sale or purchase of securities...'.(2)

Deputy Governor Harvey was called back on Day 39 of the 1929–30 Macmillan Committee, on which Keynes sat, for more discussion of monetary policy and its implementation.(3) He made an opening statement (Qu 7512) which includes the following:

> 'I said when I was here before that I regard it as the principal duty of a Central Bank to maintain the stability of the national monetary unit.... If I were asked to state in a few words what the Bank's policy has been, I should say that it has been to maintain a credit position which will afford reasonable assurance of the convertibility of the currency into gold in all circumstances, and, within the limits imposed by that objective, to adjust the price and volume of credit to the requirements of industry and trade. I should say at this stage that we regard the Bank Rate as our principal weapon for carrying that policy into effect.... In speaking of the weapons which the Bank uses I purposely omit any reference to control of the volume of credit by means of open market operations; because, after all, such operations are merely part of the machinery by which the weapon of the Bank Rate is made efficient.'

Nearly 30 years later, in his oral evidence to the Radcliffe Committee, Chief Cashier O'Brien presented a similar and a rather clear picture of the system:

> 'If the discount houses having been to all the banks and found out what they are doing, whether calling cash or lending cash, have finally come to the conclusion that the supply of cash on that day is not going to be sufficient to enable them to carry their books of Treasury Bills and short bonds, they can come to us and we can repair the shortage in one of two ways. The Chief Cashier can buy bills from the market at the market rate. There is no pain to the discount houses; they merely exchange part of their bill portfolio at the going market price for cash. That puts them square. If we are not disposed to help them in that painless way then they have to come round to the Discount Office, and nowadays borrow at Bank Rate on the security of market Treasury Bills or short bonds.'(4)
>
> '*Professor Cairncross*: ...where you are acting as lender of last resort, you are lending at Bank Rate?—Acting as lender of last resort, it is at Bank Rate. The other method, of buying in bills is not lending: it is putting out cash in exchange for securities.'(5)

Other of O'Brien's replies seem to imply clearly that the operating target was an interest rate, rather than a quantity. Moreover, they do not imply that OMOs were directly setting the market rate, but rather that OMOs were used to adjust the quantities so as to control the volume of borrowing at Bank Rate.

> 'If we wanted to raise interest rates, then we would give less help or possibly no help at all, and we would say: 'If you want cash you must come to the Discount Office for it'. And moreover we could if need be so arrange that the market needed a great deal of cash; the influence can be graded almost infinitely.'(6)

Assistant Director Coleby's 1982 paper on operational procedures for meeting monetary objectives brings out

(1) It draws on extensive research by Roger Clews, whose work almost gives the 'archaeology of knowledge' a good name.
(2) *Lord Norman* by Sir Henry Clay, London: Macmillan, 1957.
(3) Sir Ernest Harvey was only the second 'full-time central banker' to be appointed to the Bank's Court of directors, and was subsequently Deputy Governor for seven years. According to his obituarist, in his evidence to the MacMillan Committee 'his candour, no less than his grasp of the subject, greatly impressed those members of the Committee who had been most critical of the Bank's constitution and policy'.
(4) *Committee on the working of the monetary system: minutes of evidence*, London: HM Stationery Office, 1960, question 90.
(5) *Ibid*, question 93.
(6) *Ibid*, question 98.

the continuity in the Bank's operational framework until the early 1980s reforms:

> 'Until recently and despite variations and appearances to the contrary, the operational technique for giving effect to official interest rate objectives has stayed close to the classical model. That involved the setting, and periodic variation, of an official discount or lending rate, which, when necessary, is 'made effective' by open market operations in the money market. 'Making Bank rate effective' means restraining a decline in market rates from an unchanged Bank rate, or bringing them up to a newly established and higher Bank rate; it is accomplished by limiting the availability of cash to the banking system so as to 'force the market into the Bank' to borrow at the somewhat penal rate of Bank rate.'(1)

One of the 'variations' to which Coleby referred was the replacement in 1972 of Bank Rate with a Minimum Lending Rate related by formula to the result of the latest Treasury bill tender. This change was made so that the government, negotiating on pay and price controls, might avoid the accusation that it had raised the price of money.(2) This arrangement could be and was overridden—five times before it was replaced by an explicitly administered MLR in 1978. Setting a very short lending rate by reference to a three-month market rate had proved uncomfortable.

The reforms of the early 1980s came out of the debate on monetary base control (MBC) initiated by the first Thatcher government elected in 1979. Although the case for MBC was eventually rejected by the government, the new money market arrangements were designed to leave open a move in that direction and anyway to loosen official control over rates and give the market more influence on rates within an 'unpublished band'. Continuous posting of MLR was abandoned, as was the preannouncement of OMO dealing rates and the practice of deliberately creating a shortage by overissuing Treasury bills on HMG's behalf. It was also at this point that the Bank began to publish each day its estimate of the market shortage or surplus—relative to the clearing banks' desired operational balances. The Bank aimed to 'broadly offset the cash flows between the Bank and the money markets' so as to leave the clearing banks within reach of their desired balances. The aim was to do this primarily through OMOs and not through lending to the discount houses. In a 1986 BIS paper, Tony Coleby described the scheme's design in the following terms.(3)

> 'The normal conduct of the Bank's money-market operations was therefore envisaged as a market-clearing exercise. The Bank would accept as many of the offers or bids as was necessary to square the market, starting with the best rates and arriving at the 'stop rate' which just cleared the market, the result of the operation, including the range of rates at which the dealings had taken place, being immediately made public. Provision was made to override the normal arrangements if they should produce a stop rate which was unacceptable to the authorities, by lying outside an 'unpublished band' which defined the range of short-term interest rates currently judged to be consistent with policy objectives. If the stop rate was too high, more cash would be put into the system so as to arrive at an acceptably lower one: if too low, the system would be left short of cash so as to drive rates up.'

The system was clearly designed with variable-rate OMOs in mind. But with no continuously posted official rates to offer a focus to the market, the 'stop rates' accepted in the OMOs acquired great significance: '... every downward movement, even as small as $^{1}/_{16}$ per cent, came to be seen as a signal of official intent, not as a passive or incidental response to market fluctuations'.(4)

In the early days of the new regime the market did sometimes initiate rate changes (in the form of changes to banks' base rates, which the Bank then followed in its operations). Later, there were 'growing misgivings among policy-makers over the market's ability to provide a valid second opinion on the conduct of policy',(5) and the official hand was not so hidden. The 1981 regime had reserved the Bank's right to reinstate MLR intermittently by announcing the minimum rate which, for a short period ahead, would apply to any lending to

(1) Published as 'The Bank of England's operational procedures for meeting monetary objectives', *Bank of England Quarterly Bulletin*, June 1983, page 213.
(2) See Goodhart 2004, *op cit*.
(3) 'Changes in money-market instruments and procedures in the United Kingdom', in *Changes in money-market instruments and procedures: objectives and implications*, BIS, March 1986.
(4) *Ibid*.
(5) *Ibid*.

the discount houses. And from 1985 onwards, the Bank did announce MLR from time to time, with the discount houses being invited to borrow at MLR at 2.30 pm (ie the market was 'forced into the Bank'). But, crucially for this paper, the mechanics of the operations on other days remained basically unchanged. In particular, the form was still that, in its OMOs, the Bank was responding to rate offers from its counterparties, with the consequence that it was sometimes frustrated in its attempts to implement a change in rates desired by the authorities.

There were also other residual elements of 'Bank Rate' as a penal rate. Dealing rates in OMOs, when translated from discount rates to interest rates (or yields), were usually lower than the rates charged when the market was forced into the Bank. Even when inflation targeting was introduced in the early 1990s, with official rates explicitly decided by the Chancellor, the Bank's OMO dealing rates were often lower than the rate officially announced. (This was changed in the 1996–98 reforms: see Annex 1.)

One consequence of OMOs taking centre stage was that it fostered a perception that the maturity of rates being set (or targeted) for policy purposes was the same as the maturity of the OMOs. So, for example, the Bank said: '... if official operations could be confined to the shortest paper—maturities of, say, not longer than one month—it would enable the market to become the dominant influence on the shape of the yield curve for longer money-market maturities without requiring either lead or validation from the authorities'.(1) The clear implication was that conducting OMOs at a particular maturity entailed an element of setting or validating rates at that maturity.

In the event, the Bank was not able to restrict its OMOs to short maturities. For a variety of technical reasons,(2) the scale of the banking system's structural shortage and so of the Bank's OMOs increased a lot during the early 1980s. But, until the 1990s, the Bank's daily OMOs were conducted via purchases of commercial bills and Treasury bills, and in consequence there was not enough eligible paper for the Bank to recycle liquidity at short maturities. The result was that the Bank occasionally bought bills with maturities out to three months. From time to time, that fostered perceptions that the Bank was giving signals about official policy intentions.

Through the 1990s, the Bank did not analyse the instruments of monetary policy implementation very closely. Papers typically described the central bank as setting the pivotal interest rate in its role as the marginal source of funds to the economy but did not explore OMOs and standing facilities separately. The Bank's analysis was reflected in 'Monetary policy instruments: the UK experience',(3) which stated clearly that the central bank has to be the marginal source of funds while playing down the distinction between OMOs and standing facilities: 'The classical dichotomy is between open market operations, on the one hand, and discount rate or standing facilities, on the other. *A priori*... the differences between these are more apparent than real, especially when window borrowing is secured on collateral.... Consistent with the increasing market orientation of operations, there has been a gradual shift toward use of open market operations through the 1970s and 1980s, but with the discount window available as a backstop.... Since the Bank of England always stands ready to deal daily in its operations, this mid-point for short-term rates is reinforced frequently.... So there is no need for a formal band or corridor...'.

In other words, the Bank had slipped into thinking of OMOs as the instrument through which we implemented policy.

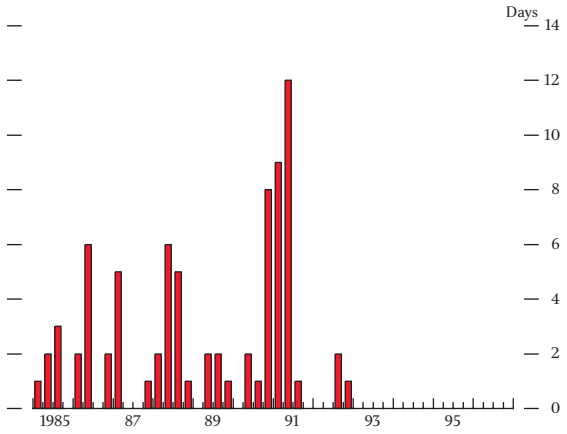
The mid-1990s reforms achieved many useful—indeed vital—improvements, but they did not include a review of the overall framework. That was because, as described in the main paper and Annex 1, they addressed urgent problems with the Bank's counterparties and with collateral. They also made important technical changes, such as conducting OMOs only at a short maturity (two weeks) and actually dealing at the announced official policy rate (ie the discount rates used for outright purchases of bills were set to produce yields at the official rate). With no deposit facility in 1997(4) and lending facilities seen as technical 'squaring up' devices (if our forecast was slightly wrong or if, because of frictions, the OMOs were not used), the OMO rate was a natural way to express policy and we slipped into thinking of it as how we actually implemented policy too. That was a fallacy, as explained in the main paper.

(1) Coleby 1986, *op cit*.
(2) See the box on 'Overfunding and money market operations', *Bank of England Quarterly Bulletin*, June 1982, page 201.
(3) *Bank of England Quarterly Bulletin*, August 1994, pages 268–276, a paper to which a wide range of Bank officials contributed.
(4) The 1996 reform proposals aired the possibility of a deposit facility to put a floor on rates. One was introduced in 2001.

The post-1996/97 system also abolished what had been known as '2.30 pm lending at Minimum Lending Rate'. It was got rid of for two reasons. One was that, apart from such lending having typically been for a maturity of a week, it seemed simply to add yet another layer of complexity in a system that in any case provided for 'late lending' to the market.(1) The other was that use of MLR was tied up with signalling (ie with 'public demonstrations' of the authorities' desired level of rates), and we had moved to a monetary regime where signalling via the Bank's operations was not needed: the official interest rate was decided at a monthly meeting (first by the Chancellor, subsequently the MPC) and simply announced. Indeed, the perception that the Bank might signal had, at times, been a complicating factor during the early 1990s. Explicitly ruling out any such possibility was one of the lasting benefits of the 1996–97 reforms. Reflecting those considerations, '2.30 pm lending' had not been used since the immediate aftermath of the ERM crisis (Chart A). So we got rid of it. Although that did not in principle impair our ability to set rates, it does nicely capture how far the Bank's analysis had drifted away from its historical and analytical base: 2.30 pm lending was the direct descendent of the apparatus used by previous generations to make 'Bank Rate effective', ie to set rates!

In fact, as recorded in the main text, the operational planning in the mid to late 1990s did reflect an understanding of the 'classical system':

Chart A
Usage of '2.30 pm lending'



> 'In addition, we may also in some conditions need to leave some of the daily shortage to be relieved at the end of the day via late lending so as to ensure that the Bank is—and is known to be—the marginal supplier of system liquidity throughout the day, preventing large banks from substituting themselves as the marginal player(s). The known availability of late lending at a known rate should also help to put a cap on the upward volatility of very short rates.'

But such discretion has not been used, in order to avoid any risk of the Bank being perceived to give signals via its operations about the MPC's rate intentions.

(1) Historically, the 'late lending' facility for the discount houses seems to have been thought of as 'banking' rather than an instrument of policy—a muddled distinction which may go back to the 1844 Act's separation of Issue Department and Banking Department.