

9. The World that Bagehot Knew

Review

Up to now, we have been looking at the money market as a mechanism for making a decentralized payments system operate as close as possible to the ideal one-big-bank system. We started with the observation that, in principle, all payments could be made as book entries on the balance sheet of a single, central bank. This is an ideal, and we can view the institutional arrangements of our actual payments system as an attempt to get as close as possible to that ideal.

From this point of view, in a decentralized payments system, the central bank emerges as a kind of clearinghouse for facilitating interbank payments by netting offsetting payments. The motivation for such an institution is to promote elasticity in the payments system, so that any two people who want to make a trade can in fact do so, while at the same time imposing discipline through, for example, periodic net settlement.

Farther down the hierarchy, we can think of the money market as a system for facilitating interbank payments by enabling surplus and deficit units to find each other easily. Focusing narrowly on banks, we saw how the Fed Funds market enables banks experiencing temporary net cash outflows to borrow needed reserve funds from the banks receiving those outflows as inflows. Thinking about security dealers, we saw how the repo market serves much the same function of allowing dealers with cash deficit to borrow from dealers with cash surplus.

We saw also how, in both cases, there is a well-developed backstop in case surplus and deficit units fail to find one another. The Fed discount window (and hence the Fed balance sheet) operates as a backstop for the Fed Funds market, and lines of credit at the dealers' clearing banks operate as a backstop for the repo market.

Outside the U.S., the same principles apply in the Eurodollar market. Because foreign banks do not have access to the Fed discount window, they have evolved different ways of making sure that cash inflows match cash outflows, mainly by matching the two in advance using Forward Rate Agreements, and then relying on the overnight Eurodollar market to handle any remaining mismatch. In a similar way, foreign banks have developed mechanisms for matching up inflows and outflows in every currency in which they deal, mainly by using forward currency markets and currency swaps.

By focusing on the payments system, we were able to see the “survival constraint” in action. From this point of view, the significant conceptual distinction was between **money** (=final settlement, good funds) and **credit** (=postpone settlement, promise to pay). If you cannot meet your obligations at the final settlement (clearing) then you must postpone (borrow) or else fail (not survive). Whatever level in the system we are looking at, a payments perspective suggests that the difference between money and credit is all about **time**.

A key feature of the system is its hierarchical character, insofar as what counts as money and what counts as credit depends on what level of the hierarchical system we are considering. That is why the next higher level can operate as a backstop for credit markets at every level. The availability of that backstop is the place where the balance between elasticity and discipline is managed. Smooth operation of the payments system requires elasticity, so that shortage of means of payment does not get in the way of mutually beneficial trade. But it also requires discipline, so final settlement is only postponed, not put off forever.

We see the tension between these two principles in the simultaneous elasticity of credit and inelasticity of money. We see the tension also in institutions like continuous real time gross settlement and control over daylight overdrafts on Fedwire, limits on overdrafts and daily net settlement on CHIPS, combined with looser limits in the repo and Eurodollar markets. The effect is to protect the core of the payments system from potential credit problems (excess elasticity), and to protect the core of the credit system from potential money shortage problems (excess discipline).

In terms of our four prices of money, we have so far been emphasizing one, par, which is the price of deposit money in terms of currency (or reserve money). We have referred to the interest rate as the price of delaying settlement, but we haven't talked very much about what determines that rate. In our discussion, the clearinghouse was presumed to receive no interest on its assets (gold) and pay none on its standing liabilities (CH certificates). In times of crisis, the clearinghouse was presumed to charge an exceptionally high rate of interest (6%) on loans to members, and to pass that high rate of interest on to members who were willing to hold clearinghouse loan certificates. We see here a primitive kind of lender of last resort function emerging organically from business practice. But such a rate of interest bears no close relation to the rate of interest in normal times.

Preview

The next section of the course approaches the money market from a new point of view. Instead of thinking of credit instruments as forms of delayed payment (different times), we think of them as securities that have a current **price** quoted in terms of money (different prices). The price we will focus most of our attention on is the rate of interest, which is the price of money today in terms of money tomorrow. Furthermore, instead of focusing on the various credit instruments themselves (Fed Funds, repo, Eurodollars), we focus now on the institutions that make "liquid" markets in these instruments. The most significant idea will be that **banks are a kind of security dealer, a market maker in money**. We won't be able to understand that idea fully until next lecture, since first we have to understand what is meant by a security dealer, but it will help you to understand where we are going with this.

In a way, we are reversing the emphasis of Stigum. She says: "the dealer takes in securities on one side at one rate and hangs them out on the other side at a slightly more favorable (lower) rate; or to put it the other way around, the dealer borrows money from his repo customers at one rate and lends it to his reverse customers as a slightly higher rate. In doing so, the dealer is acting like a bank, and dealers know this well" (432). She emphasizes that dealers are like banks, meaning they are a kind of financial intermediary between ultimate borrowers and lenders. I want to emphasize that banks are like dealers, meaning a kind of market maker. And not only are individual banks like dealers, but also the central bank is like a dealer.

These hierarchical layers of market making knit together the layers of the hierarchy of money. The ordinary market participant's impression that the difference between instruments is a matter of quantity (price) not quality (moneyness) is created by this system of market makers. The hierarchy is still there behind the scenes, and occasionally springs out onto the stage during crises, but most of the time we don't see it.

In this next section of the course, we will be focusing on the money rate of interest as the price of money. This new focus means that we now approach the emergence of central banking also from another angle. We think of banks as directly managing the relationship between payment and promises to pay, which is to say the relationship between money today and money tomorrow, which is to say the rate of interest. Just as in the last few lectures we pursued an analogy between a central bank and a clearinghouse, now we pursue a different analogy between **a central bank and a security dealer**. We begin with 19th century British banking where that analogy can be seen in a simpler form than in modern banking.

The Discount Mechanism

Let us suppose that firm A buys goods wholesale from Firm B. Instead of paying immediately, he promises to pay in 30 days. The tangible evidence of such a promise is a “bill of exchange”. Let us suppose further that firm B does not wish to hold the bill, because he needs to make payments of his own (perhaps to his workers, or to repay maturing bills). So he takes the bill to a bank which “discounts” it, paying out notes less than the full face value of the bill.¹ The difference is interest which the bank will earn by holding the bill to maturity. Let us suppose further that the bank buys the bill with notes that it was holding in reserve, so that its reserves fall. Then we have the following:

Firm A		Firm B		Bank	
Assets	Liabilities	Assets	Liabilities	Assets	Liabilities
+goods	+bill of exch.	-goods +bill of exch.			
		-bill of exch. +notes		+bill of exch. -notes	

Note what has happened. The balance sheet of Firm A has expanded on both sides, while Firm B and the Bank have merely changed the form in which they hold their assets. If all goes well, Firm A manages to sell the goods retail, receiving notes in return which he then uses to pay back what he owes to the bank, so shrinking his balance sheet again.

Firm A		Firm B		Bank	
-goods +notes					
-notes	-bill of exch.			-bill of exch. +notes	

¹ “discount” means sale at a price less than face value. Historically banks also “accepted” bills without discounting them, which means they guaranteed payment. This acceptance is a kind of credit default swap, which we will talk about in a later lecture.

So the balance sheet of Firm A shrinks back down again, and the note reserve of the bank is restored. (For simplicity in the balance sheets, I abstract from any profit firm A might have earned, as well as from the interest the bank has earned by holding the bill to maturity. In fact the notes paid in are more than the notes paid out, by the rate of interest.) Firm A financed its purchase by borrowing first from Firm B, who passed on the promise to the Bank. The result was that Firm B got paid for the goods before Firm A made the payment.

The way we did the balance sheets, that payment came from notes that the Bank held in reserve. I did it the way I did in order to make clear the cost to the Bank of helping Firm B receive a payment before Firm A makes a payment, namely a reduction in reserves relative to total assets, and hence increased vulnerability to demands for payment. The bank opens itself to that vulnerability in order to earn interest on the bill it discounts. This vulnerability is what we mean by liquidity risk. Concretely, with fewer notes on hand, the bank is more vulnerable to an unexpected withdrawal of its deposit liabilities.

Alternatively, the bank could discount by expanding its deposits. This too increases liquidity vulnerability since the new deposits may well be withdrawn, but probably less so than paying out notes right away.

Firm A		Firm B		Bank	
A	L	A	L	A	L
		-bill		+bill	
		+deposit			+deposit

One other possibility is also worth noting, which is that the firm does not discount the bill but rather only “accepts” it, which means committing to pay the bill at maturity in case Firm A does not pay. Acceptance is thus a kind of guarantee, a contingent liability of the Bank and a contingent asset of Firm B. Firm B pays for this guarantee, but in return gets an interest bearing asset that he can more readily liquidate if it turns out he needs cash before maturity.

Firm A		Firm B		Bank	
A	L	A	L	A	L
		+acceptance			+acceptance

Both of these mechanisms operate to economize on scarce notes (money). In the first, the bank finds a way to discount without paying out notes. In the second, the Firm finds a way to ensure access to future cash without demanding any present cash.

The Discount Rate

Now suppose that the bank is doing business not just with Firm A and B but with lots of other firms, and suppose there are lots of other banks too. On any given day it is both extending new

discount loans, and receiving payment from maturing discount loans, so notes are flowing in and out. The whole secret of successful banking of this type is to match these inflows and outflows while keeping (non-earning) note balances as low as possible.

In order to do this, the one thing the bank can control is the discount rate. If it finds that requests for discount are exceeding its available note capacity, it can lower the price it is willing to pay (which means raising the discount interest rate). The effect should be to discourage new discounts relative to repayment of old ones, and hence a shift toward net inflow of notes.

$r \uparrow$ discourages discount, contraction of credit at given bank

$r \downarrow$ encourages discount, expansion of credit at given bank

Every bank is doing this. Each one is moving its own discount rate up and down in order to achieve the desired net flow of paying assets. If any bank finds itself with too many notes, it may lower the discount rate in order to attract new discounts. If it finds itself with too few notes, it may raise the discount rate in order to discourage new discounts. In the system as a whole multiple rates encourage holders of bills to seek out one bank or another.

Also, banks themselves may borrow and lend notes to and from one another by using rediscount. A bank finding itself with too many discounts and too few notes can rediscount some of the excess to another bank in the opposite position. The significant point about this, for our purposes, is that the discount banks are acting in effect as dealers, quoting a two-way market for bills of exchange and meeting the resulting flow of demand and supply by altering their own inventories and balance sheets. Individual dealers set their own prices, with a view to managing their own inventories, maximizing profits and controlling exposure to liquidity risk. But competition and **arbitrage ensures that the multiple rates are coordinated around a single rate we'll call the market rate of interest.**

Given what we've said about how banks adjust their discount rates, it follows that the marketwide rate of interest fluctuates depending on the marketwide balance between cash inflows from maturing bills and cash outflows from new discounts. When discount pressure is high, or when firms are having difficulty repaying, the market rate of interest rises.

Bank Rate

Now let us think about the role of the central bank. Ever since Peel's Act, the Bank of England was divided into two departments:

Bank of England

Assets	Liabilities
Issue Department Gold Government Debt	Notes
Banking Department Notes Discounts Advances	Deposits

The effect of this division was to fix the note issue in aggregate, which is a source of discipline. The elasticity in the system comes from the willingness of the Bank of England to hold more note reserves than it really needs, and to make them available to other banks through the discount mechanism. A further source is the elasticity of deposits, which other banks are able to use as reserves when notes are particularly scarce.

So far we have been implicitly thinking of the banking problem as merely involving allocation of a fixed quantity of notes. Fluctuating market rates create a profit incentive for the banking system to expand and contract its overall balance sheet, by holding fewer notes when credit demand is higher. But it also creates a profit incentive for whoever issues the notes to change the quantity. So a profit-seeking central bank would respond to a period of general credit expansion as follows:

Banking System		Central Bank	
Assets	Liabilities	Assets	Liabilities
-bills +notes +deposits		+bills -notes	+deposits

Just as the individual banks were encouraged to take on liquidity risk by the prospect of profit, so too would be a profit-seeking central bank. As in the system with a fixed quantity of notes, there will be a tendency for all discount rates to converge on a single rate.

But notice that whenever there is scarcity of note reserve (i.e. “money is tight”), the central bank is in a position to affect the interest rate as a matter of its own policy. It may take the view that there are sufficient notes outstanding already, and set its rate sufficiently high that net demand for rediscount is minimal. Or it may set a lower rate, and allow its balance sheet to expand. In fact it may use the rediscount rate to control the rate of that expansion. Contrariwise, it may take the view that there are excessive notes outstanding, and set its rate so high that net demand for rediscount is negative, so contracting its balance sheet, and it may use its discount rate to control the rate of contraction.

Bank rate thus moves as a policy variable, while the market rate moves as a market price.

The Art of Central Banking

Here we have the origin of monetary policy. The question of course is how the central bank decides what is the right rate of discount. The beginning of wisdom, in this respect, is the realization that narrow profit-maximization is not obviously the right policy for the central bank. Rather, its ability to relax the survival constraint for banks lower down in the hierarchy brings with it responsibility for managing the balance between discipline and elasticity, choosing when to relax the constraint and also when to tighten it. The point to hold on to here is that a central bank is at a level above other banks.

In times of crisis, bills are not repaid on time so banks face a reduction in cash inflow. Unless they can replace that cash inflow by discounting at the central bank, they will be forced (the survival constraint) to curtail their own discounts, and the result will be a general credit contraction and perhaps a cascade of payment failures. Against that possibility, the central bank holds excess reserves for the system as a whole, so that in times of crisis it can “lend freely but at a high rate”. This is the Bagehot Principle established by Walter Bagehot in his Lombard Street (1873). (Bagehot called this “lender of last resort”, but I will call it “dealer of last resort” in order to fix ideas that will be more useful when we extend our analysis to modern conditions.)

In the face of an **internal drain**, the job of the central bank is to lend out its own note reserves “to the last farthing”, and even possibly to create additional deposits as a substitute for notes. The reason for the “high rate” is to make sure that banks asking for accommodation really need it, and to provide incentive for them to pay back as soon as conditions return to normal. See the balance between elasticity (lending freely) and discipline (high rate)?

The central bank can help with an internal drain because it stands above other banks, but the same is not true of an **external drain**. The central bank is not above other central banks. Suppose the demand for discounts at the central bank is coming from people who really want gold, the international money, so they convert the notes they receive into gold and take gold out of the country. Then the central bank has a problem that it cannot fix by paying out freely, much less by expanding its own balance sheet. People don’t want notes, and they don’t want deposits, they want gold, the international reserve.

Young is at great pains to describe the banking department of the Bank of England as essentially a credit operation built on an extremely small foundation of gold reserves, so that any incipient gold outflow calls for immediate action. Because of the enormous volume in the bill market, there are always bills maturing and new bills being offered. All one has to do to forestall a gold drain is to make new borrowing slightly less attractive by raising the rate of interest, which is the same thing as lowering the price of bills. Then maturing bills are not replaced but rather paid and the resulting gold inflow stems the tide.

Here we see the central bank operating according to the very same principles we laid out for banks lower in the system:

$r \uparrow$ discourages discount, contraction of credit at given bank
 $r \downarrow$ encourages discount, expansion of credit at given bank

This is the economics of Walter Bagehot, focusing on the discount rate as defense of the rate of exchange under the gold standard. When the Bank of England tightens, it contracts not only its own balance sheet but also the balance sheets of all other banks below it, which tightens credit in the entire country. Thus, to some extent, it could be said that domestic credit policy was subordinated to the dictates of the gold standard, at least whenever those dictates were binding.

This distinction between external drain and internal drain is thus an early attempt to find the appropriate balance between discipline (external) and elasticity (internal). The distinction makes a certain practical sense in a gold standard world since if the central bank does not defend the gold parity it is not clear who else would. But from a larger analytical perspective, the distinction is a lot muddier than it appears on first sight. If we take the perspective of the world as a whole, then gold flows between countries have no more effect on the aggregate world gold stock than internal flows have on the domestic gold stock. If we accommodate the latter, then why don't we accommodate the former?

Practically of course the reason is that such accommodation would require cooperation between central banks—the French central bank would have to lend back the gold that was flowing out of England. But that is only a practical problem, and it is conceivable that evolution of central banking would provide an institutional solution. There is no necessity for the scarcity of gold to constrain the system as a whole. The system does however require some overall disciplining factor.