MILTON FRIEDMAN

MONEY MISCHIEF

Episodes in Monetary History

HARCOURT BRACE JOVANOVICH, PUBLISHERS
New York San Diego London

CHAPTER 2

The Mystery of Money

The term *money* has two very different meanings in popular discourse. We often speak of someone "making money," when we really mean that he or she is receiving an income. We do not mean that he or she has a printing press in the basement churning out greenbacked pieces of paper. In this use, money is a synonym for income or receipts; it refers to a flow, to income or receipts per week or per year. We also speak of someone's having money in his or her pocket or in a safe-deposit box or on deposit at a bank. In that use, money refers to an asset, a component of one's total wealth. Put differently, the first use refers to an item on a profit-and-loss statement, the second to an item on a balance sheet.

In this book I shall try to use the term *money* exclusively in the second sense, as referring to an item on a balance sheet. I say "try" because, with use of the term as a synonym for income or receipts so ubiquitous, I cannot guarantee that even I, who have been aware for decades of the importance of distinguishing between the two uses, will not occasionally slip and use the term in the first sense.

One reason why money is a mystery to so many is the

role of myth or fiction or convention. I started this book with the chapter on stone money precisely in order to illustrate that point. To make the same point in a way that is perhaps more relevant to the everyday experience of most of us, consider two rectangles of paper of about the same size. One rectangle is mostly green on the back side and has a picture of Abraham Lincoln on the front side, which also has the number 5 in each corner and contains some printing. One can exchange pieces of this paper for a certain quantity of food, clothing, or other goods. People willingly make such trades.

The second piece of paper, perhaps cut from a glossy magazine, may also have a picture, some numbers, and a bit of printing on its face. It may also be colored green on the back. Yet it is fit only for lighting the fire.

Whence the difference? The printing on the five-dollar bill gives no answer. It simply says, "FEDERAL RESERVE NOTE / THE UNITED STATES OF AMERICA / FIVE DOLLARS" and, in smaller print, "THIS NOTE IS LEGAL TENDER FOR ALL DEBTS, PUBLIC AND PRIVATE." Not very many years ago, the words "WILL PROMISE TO PAY" were included between "THE UNITED STATES OF AMERICA" and "FIVE DOLLARS." Did that mean the government would give you something tangible for the paper? No, it meant only that if you had gone to a Federal Reserve bank and asked a teller to redeem the promise, the teller would have given you five identical pieces of paper having the number 1 in place of the number 5 and George Washington's picture in place of Abraham Lincoln's. If you had then asked the teller to pay the $1.00 promised by one of these pieces of paper, he would have given you coins, which, if you had melted them down (despite its being illegal to do so), would have sold for less than $1.00 as metal. The present wording—no

longer with a "promise to pay"—is at least more candid, if equally unrevealing.

The "legal tender" quality means that the government will accept the pieces of paper in discharge of debts and taxes due to itself and that the courts will regard them as discharging any debts stated in dollars. Why should they also be accepted by private persons in private transactions in exchange for goods and services?

The short answer—and the right answer—is that private persons accept these pieces of paper because they are confident that others will. The pieces of green paper have value because everybody thinks they have value. Everybody thinks they have value because in everybody's experience they have had value—as is equally true for the stone money of chapter 1. The United States could barely operate without a common and widely accepted medium of exchange (or at most a small number of such media); yet the existence of a common and widely accepted medium of exchange rests on a convention: our whole monetary system owes its existence to the mutual acceptance of what, from one point of view, is no more than a fiction.

That fiction is no fragile thing. On the contrary, the value of having a common money is so great that people will stick to the fiction even under extreme provocation. But neither is the fiction indestructible: the phrase "not worth a Continental" is a reminder of how the fiction was destroyed by the excessive amount of Continental currency the Continental Congress issued to finance the American Revolution.

The numerous inflations throughout history—whether the recent moderate inflations in the United States, Britain, and other advanced countries, or the very large recent inflations in South and Central American countries, or the

hyperinflations after World Wars I and II, or the more ancient inflations going back to Roman times—have demonstrated the strength of the fiction and, indirectly, its usefulness. It takes very high rates of inflation—rates well up in double digits that persist for years—before people will stop using the money that is so obviously inflating. And when they do lose faith in the fiction, they do not revert to straight barter. No, they adopt substitute currencies. The substitute currencies in most inflations in history have been gold, silver, or copper specie, often, as during the American Revolution, in the form of coins of foreign countries. What's more, people may not abandon paper altogether: they may turn instead to paper money that has not been overissued.

Two particularly revealing examples are provided by the American Revolution, more than two centuries ago, and the Russian Revolution of 1918. The Continental currency was overissued. The result was that the promise of redemption in specie was not honored, and Continental currency came to be accepted only at the point of a gun. On the other hand, some of the original thirteen colonies issued their own paper money, which remained limited in amount, and this paper money continued to be used, along with coins of foreign countries. An even more striking example is that provided by the Russian Revolution of 1918, which was followed by a hyperinflation of far greater magnitude than the American revolutionary inflation. When, in 1924, the inflation was ended and a new currency established, one of the new chervonets rubles was exchanged for 50 billion old rubles! *These old rubles were the ones that had been issued by the new Soviet government.* There also still existed old czarist paper rubles. Since there was small prospect that a czar would return to redeem the

promise printed on the czarist rubles, it is remarkable that they were still being accepted as substitute currency and had retained their purchasing power. They retained their value precisely because no new czarist rubles could be created, and hence the quantity available to circulate was fixed.

During the German hyperinflation after World War I, currencies of foreign countries served as a substitute currency. After World War II, the Allied occupational authorities exercised sufficiently rigid control over monetary matters, in the course of trying to enforce price and wage controls, that it was difficult to use foreign currency. Nonetheless, the pressure for a substitute currency was so great that cigarettes and cognac emerged as substitute currencies and attained an economic value far in excess of their value purely as goods to be consumed.

I personally experienced a remnant of the use of cigarettes as money in 1950, by which time monetary stability had been restored to Germany and the paper German mark was again the common medium of circulation. Driving from Paris, where I was spending a few months as a consultant to the U.S. agency administering the Marshall Plan, to Frankfurt, the newly established temporary capital of Germany and also the base of U.S. occupation authorities, I had to refill the gasoline tank of the "Quatre Chevaux" (a small Renault car) that I was driving. As it happened, I had no marks with me, because I was to get an allotment of them when I arrived in Frankfurt. But I did have dollars, French francs, and British pounds. The German *frau* who filled my tank would accept none of these in payment—that was illegal, she said. "*Haben sie keine wäre* ["Have you any goods"]?" was her next remark. We settled amicably when I gave her a carton of cigarettes (for

which I had paid $1.00 at the Paris PX—remember, this was a long time ago) for gasoline that she valued at $4.00 at the official exchange rate for marks but that I could purchase at a U.S. PX for $1.00. As she viewed it, she got $4.00 worth of cigarettes in return for $4.00 worth of gasoline. As I viewed it, I got $1.00 worth of gasoline in return for $1.00 worth of cigarettes. And both of us were happy. But, as I used to ask my students, what became of the missing $3.00?

I should add that a few years earlier, before Ludwig Erhard's 1948 monetary reform—the first step in the remarkable postwar recovery of Germany—a carton of cigarettes would have been valued as the equivalent of a far larger number of marks than the number that, at the then official exchange rate, could have been purchased for $4.00. As currency, cigarettes were typically traded by the pack, or even the single cigarette, not by the carton—that would have been far too high a denomination for most purchases. Foreigners often expressed surprise that Germans were so addicted to American cigarettes that they would pay a fantastic price for them. The usual reply was: "Those aren't for smoking; they're for trading."

As the example of cigarettes (or cognac) suggests, an amazing variety of items have been used as money at one time or another. The word "pecuniary" comes from the Latin *pecus*, meaning "cattle," one of the many things that have been used as money. Others include salt, silk, furs, dried fish, tobacco, even feathers and, as we saw in chapter 1, stones. Beads and cowrie and other shells, such as the American Indians' wampum, have been the most widely used forms of primitive money. Metals—gold, silver, copper, iron, tin—have been the most widely used forms among advanced countries before the victory of paper and

the bookkeeper's pen (although a temporary use of paper as money occurred in China more than a millennium ago).

What determines the particular item that will be used as money? We have no satisfactory general answer to that simple question. We do know that, however the habit of using one item or another as money arises, the habit takes on a life of its own and, like Topsy, just grows. As Walter Bagehot, a nineteenth-century editor of the English periodical *The Economist* puts it in his masterpiece, *Lombard Street:* "Credit is a power which may grow, but cannot be constructed" (1873, p. 69). Substitute "unit of account" or "money" for "credit" to translate that statement into the terms we have been using.

We can come closer to giving a reasonably general answer to a different, and basically more important, question: What determines the value in terms of goods and services of whatever item has come to be accepted as money?

When most money consisted of silver or gold or some other item that had a nonmonetary use, or of an enforceable promise to pay a specified amount of such an item, the "metallist" fallacy arose that "it is logically essential for money to consist of, or be 'covered' by, some commodity so that the logical source of the exchange value or purchasing power of money is the exchange value or purchasing power of that commodity, considered independently of its monetary role" (Schumpeter 1954, p. 288). The examples of the stone money of Yap, of cigarettes in Germany after World War II, and of paper money currently make clear that this "metallist" view is a fallacy. The usefulness of items for consumption or other nonmonetary purposes may have played a role in their acquiring the status of money (though the example of the stone money of Yap

indicates that this has not always been the case). But once they acquired the status of money, other factors clearly affected their exchange value. The nonmonetary value of an item is never a fixed magnitude. The number of bushels of wheat or pairs of shoes or hours of labor that an ounce of gold can be exchanged for is not a constant fact. It depends on tastes and preferences and on relative quantities. The use of, say, gold as money tends to alter the quantity of gold available for other purposes and in that way to alter the amount of goods that an ounce of gold can be exchanged for. As we shall see in chapter 3, in which we analyze the effect of the demonetization of silver in the United States in 1873, the nonmonetary demand for an item used as money has an important effect on its monetary value, but, similarly, the monetary demand affects its nonmonetary value.

For present purposes, we can simplify our attempt to demystify money by concentrating on the monetary arrangement that, while historically a very special case, is currently the general rule: a pure paper money that has practically no value as a commodity in itself. Such an arrangement has been the general rule only since President Richard M. Nixon "closed the gold window" on August 15, 1971—that is, terminated the obligation that the United States had assumed at Bretton Woods to convert dollars held by foreign monetary authorities into gold at the fixed price of $35 an ounce.

Before 1971, every major currency from time immemorial had been linked directly or indirectly to a commodity. Occasional departures from a fixed link did occur but, generally, only at times of crisis. As Irving Fisher wrote in 1911, in evaluating past experience with such episodes: "Irredeemable paper money has almost

invariably proved a curse to the country employing it" (1929, p. 131). As a result, such episodes were both expected to be and were temporary. The link was successively weakened, however, until it was finally eliminated by President Nixon's action. Since then, no major currency has had any link to a commodity. Central banks, including the U.S. Federal Reserve System, still carry an entry on their balance sheets for gold, valued at a fixed nominal price, but that is simply the smile of a vanished Cheshire cat.

What, then, determines how much one can buy with the greenbacked five-dollar paper bill we started with? As with every price, the determinant is supply and demand. But that only begs the question. For a full answer, we must ask: What determines the supply of money? And what determines the demand for money? And what, concretely, is "money"?

The abstract concept of money is clear: money is whatever is generally accepted in exchange for goods and services—accepted not as an object to be consumed but as an object that represents a temporary abode of purchasing power to be used for buying still other goods and services. The empirical counterpart of this abstract concept is far less clear. For centuries, when gold and silver were the major mediums of exchange, economists and others regarded only coins as money. Later they added bank notes redeemable on demand for gold or silver specie. Still later, a little over a century ago, they accepted bank deposits payable on demand and transferable by check. Currently, in the United States, a number of monetary aggregates are regularly compiled, each of which may be regarded as the empirical counterpart of money. These range from currency, the narrowest total, to the total of specified liquid

assets, the aggregate designated "L" by the Federal Reserve.*

We can bypass this highly technical issue by considering a hypothetical world in which the only medium of circulation is paper money like our five-dollar bill. For consistency with the present situation, we shall assume that the number of dollars of such money in circulation is determined by a governmental monetary authority (in the United States, the Federal Reserve System).

The Supply of Money

Analysis of the supply of money, and in particular of changes in the supply of money, is simple in principle but extremely complex in practice, both in our hypothetical world and in the current real world. Simple in principle, because the supply of money is whatever the monetary authorities make it; complex in practice, because the decisions of the monetary authorities depend on numerous factors. These include the bureaucratic needs of the authorities, the personal beliefs and values of the persons in charge, current or presumed developments in the economy, the political pressures to which the authorities are subject, and so on in endless detail. Such is the situation that prevails today. Historically, of course, the situation was very different because the commitment to redeem government- or bank-issued money in specie meant that the physical conditions of production played a significant role. Later chapters explore the consequences of the commitment to redeem in considerable detail.

*For a full discussion of the definition of money, see Friedman and Schwartz (1970, part 1).

It's simple to state how the money supply is so centrally controlled. It's hard to believe. I have observed that noneconomists find it almost impossible to believe that twelve people out of nineteen—none of whom have been elected by the public—sitting around a table in a magnificent Greek temple on Constitution Avenue in Washington have the awesome legal power to double or to halve the total quantity of money in the country. How they use that power depends on all the complex pressures listed in the previous paragraph. But that does not alter the fact that they and they alone have the arbitrary power to determine the quantity of what economists call base or high-powered money—currency plus the deposits of banks at the Federal Reserve banks, or currency plus bank reserves. And the entire structure of liquid assets, including bank deposits, money-market funds, bonds, and so on, constitutes an inverted pyramid resting on the quantity of high-powered money at the apex and dependent on it.

Who are these nineteen people? They are seven members of the Board of Governors of the Federal Reserve System, appointed by the president of the United States for fourteen-year nonrenewable terms, and the presidents of the twelve Federal Reserve banks, appointed by their separate boards of directors, subject to the veto of the Board of Governors. These nineteen constitute the Open Market Committee of the Federal Reserve System, though only five of the bank presidents have a vote at any one time (in order to assure that the seven members of the central board have ultimate authority).

The exercise of this arbitrary power has sometimes been beneficial. However, in my view, it has more often been harmful. The Federal Reserve System, authorized by the Congress in 1913 and beginning operations in 1914,

presided over the more than doubling of prices that occurred during and after World War I. Its overreaction produced the subsequent sharp depression of 1920–21. After a brief interval of relative stability in the 1920s, its actions significantly intensified and lengthened the great contraction of 1929–33. More recently, the Fed was responsible for the accelerating inflation of the 1970s—to cite just a few examples of how its powers have in fact been used.*

The Demand for Money

The Federal Reserve can determine the quantity of money—the number of dollars in the hands of the public. But what makes the public willing to hold just that amount, neither more nor less? For an answer, it is crucial to distinguish between the nominal quantity of money—the number of dollars—and the real quantity of money—the amount of goods and services that the nominal quantity will purchase. The Fed can determine the first; the public determines the second, via its demand for money.

*For a full discussion, see Friedman and Schwartz (1963).

A somewhat amusing example of the kind of petty personal concerns that can enter into the activities of such an august body is the renaming of the ruling body in Washington in 1935. The name was changed from "Federal Reserve Board" to "Board of Governors of the Federal Reserve System." Why substitute the more cumbersome name for the more compact? The reason was entirely considerations of prestige. In central bank history, the head of a central bank has been called the governor. That is the prestigious title. Before 1929, the heads of the twelve separate Federal Reserve banks were designated as governors, in line with the desire of the founders to have a truly regional, decentralized system. On the Federal Reserve Board, only the chairman was designated governor; the other six members were simply that—members of the Federal Reserve Board. As part of the Banking Act of 1935, the heads of the separate Federal Reserve banks were redesignated presidents, and the central body was renamed the Board of Governors of the Federal Reserve System so that each member of it could be a governor! Petty, but also a symbol of a real transfer of power from the separate banks to Washington.

There are many ways to express the real quantity of money. One particularly meaningful way is in terms of the flow of income to which the cash balances correspond. Consider an individual receiving an income of, say, $20,000 a year. If that individual on the average holds $2,000 in cash, his cash balances are the equivalent of one-tenth of a year's income, or 5.2 weeks of income: his cash balances give him command over the quantity of goods and services that he can buy with 5.2 weeks' income.

Income is a flow; it is measured as dollars per unit of time. The quantity of money is a stock, not in the sense of equity traded on an exchange but in the sense of a store of goods or inventory, by contrast with a flow. Nominal cash balances are measured as dollars at a point in time—$2,000 at 4:00 P.M. on July 31, 1990. Real cash balances, as just defined, are measured in units of time, like 5.2 weeks; "dollars" do not enter in.*

It is natural for you, as a holder of money, to believe that what matters is the number of dollars you hold—your nominal cash balances. But that is only because you take dollar prices for granted, both the prices that determine your income and the prices of the things you buy. I believe that on reflection you will agree that what *really* matters is your real cash balances—what the nominal balances will buy. For example, if we expressed nominal magnitudes in cents instead of dollars, both nominal income and nominal

*The more usual practice is to define real cash balances by dividing nominal balances by a price index. The price index itself typically represents the estimated cost at various dates of a standard basket of goods (so many loaves of bread, pounds of butter, pairs of shoes, and so on, to encompass, for a consumer price index, the typical budget of a consumer). Under this definition, real cash balances have the dimension of the number of baskets of goods the nominal balances could purchase.

cash balances would be multiplied by 100, but real balances would be unaffected, and it would make no difference to anyone (except those who had to write down the larger numbers).

Similarly, try to conceive of every price, including those that determine your income, being multiplied by 100 overnight—or divided by 100—and, correspondingly, your cash balances, nominal debts, and nominal assets being simultaneously multiplied or divided by 100. Nothing would be really changed. Of course, that is not the way changes in the quantity of money or in prices generally come about, which is what raises all the difficulties in monetary analysis. But it is what happens when a government, typically during or after a major inflation, announces a so-called monetary reform that substitutes one monetary unit for another. It is what, for example, General Charles de Gaulle did in France on January 1, 1960, when he replaced the then franc with the *nouveau franc,* or new franc, by simply striking two zeros from all calculations in the old franc. In other words, 1 new franc equaled 100 old francs. De Gaulle made this change as part of an extensive monetary and fiscal reform that did have significant effects, though the mere change of units did not. However, the episode is another instance of how deeply embedded are public attitudes to money. For decades thereafter, many French residents continued to express prices and perform monetary calculations in old francs, striking off the final two zeros only when offered payment in *nouveau francs.*

When such alterations of monetary units are combined with superficial monetary and fiscal changes, as when Argentina in 1985 replaced the peseta with the austral, they have at most had highly temporary and minor effects, because they do not by themselves alter real magnitudes.

Given that it is the real quantity of money, not the nominal quantity, that matters, what determines whether people will want to hold cash balances averaging about five weeks' income—as in practice they have done in many countries over long periods of time—or only about three or four days' income—as they did, for example, in Chile in 1975?*

Two major forces determine how much cash people will want to hold: (1) the usefulness of cash balances as a temporary abode of purchasing power; (2) the cost of holding cash balances.

(1) *Usefulness.* Cash balances are useful as a means of enabling an individual to separate the act of purchase from the act of sale. In a world without money, transactions would have to take the form of barter. You have A to sell and want to acquire B. To do so you must find someone who has B to sell and wants A and must then make a mutually acceptable deal—what the textbooks dub "the double coincidence of barter." In a money economy, you can sell A for money, or generalized purchasing power, to anyone who wants A and has the purchasing power. You can in turn buy B for money from anyone who has B for sale, regardless of what the seller of B in turn wants to buy. This separation of the act of sale from the act of purchase is the fundamental productive function of money.

A related reason for holding money is as a reserve for future emergencies. Money is only one of many assets that can serve this function, but for some people at some times it may be the preferred asset.

*These illustrative numbers are chosen to correspond with the identification of money as currency and are realistic for currency. For broader aggregates, such as the U.S. M2, cash balance holdings are much larger. In the United States currently, they are about nine months of national income.

How useful money is for these purposes depends on many factors. For example, in an underdeveloped economy consisting of largely self-sufficient households, each producing mostly for its own consumption, monetary transactions are relatively unimportant. As such societies develop and the range of monetary transactions increases, cash balances rise much faster than income, so that real cash balances, expressed as weeks of income, increase. Such development generally occurs along with urbanization, which has much the same effect, because it means that a larger fraction of transactions are impersonal. Credit at the local grocery store is not likely to be as readily available to smooth over discrepancies between receipts and expenditures.

At the other extreme, in financially advanced and complex societies, such as the United States today, a wide array of assets is available that can serve as more or less convenient temporary abodes of purchasing power. These range from cash in pocket, to deposits in banks transferable by generally accepted check, to money-market funds, credit-card accounts, short-term securities, and so on, in bewildering variety. They reduce the demand for real cash balances narrowly defined, such as currency, but they may increase the demand for real cash balances more broadly defined by making temporary abodes of purchasing power useful in facilitating shifts between various assets and liabilities.*

(2) *Cost.* Cash balances are an asset and, as such, an alternative to other kinds of assets, ranging from other

*As it happens, for most of the post–World War II period these further effects have roughly balanced in the United States, so that M2, as currently defined, has ranged around nine months, mostly in response to changes in the cost of holding cash balances.

nominal assets, such as mortgages, savings accounts, short-term securities, and bonds, to physical assets, such as land, houses, machines, or inventories of goods, which may be owned either directly or indirectly, via equities, or common stocks. Accumulating an asset requires saving, that is, abstaining from consumption. Once the asset is accumulated, it may cost something to maintain, as with physical inventories, or it may yield a return in the form of a flow of income, such as interest on a mortgage or bond or dividends on stock.

As with cash balances, it is important to distinguish between the nominal return on an asset and the real return. For example, if you receive 10 cents per dollar on a bond when prices in general are rising by, say, 6 percent a year, the real yield is only 4 cents per dollar because you must reinvest 6 cents per dollar to have the same purchasing power invested at the end of the year as at the beginning.* The nominal yield is 10 percent, the real yield 4 percent. Similarly, if prices are falling, the real yield will exceed the nominal yield by the rate of the price drop.

What matters ultimately is the real, not the nominal, magnitudes. As a result, the nominal yield on assets such as bonds has tended to adjust *over long periods* in order to keep the real yield roughly the same. However, that has been very far from the case over short periods, because of the time it takes for people to adapt to changed circumstances.

For cash balances, one cost—the cost that has been stressed in monetary literature—is the interest return that is sacrificed by holding cash rather than "safe"

*In practice, the situation is more complex because of the need to allow for tax effects.

interest-earning assets, for example, the interest received per dollar of a U.S. Treasury bill as against the interest, if any, received per dollar of cash balances (zero for currency).

Another cost or return—and one less stressed, though often far more important—is the change in the real value of a dollar. If prices are rising at a rate of 6 percent a year, say, $1.00 will be able to buy only as much at the end of the year as 94 cents would have bought at the beginning. To keep the real value of your cash balances constant, you would have to hold balances of 6 percent more at the end of the year than at the beginning. On the average, it would cost you 6 cents for every dollar that you held during the year. Conversely, if prices were falling at the rate of 6 percent a year, you would in effect receive a return of 6 cents for every dollar you held during the year. Clearly, cash is a less attractive asset when prices are rising than when they are falling.

For a nominal asset, the nominal interest sacrificed and the change in purchasing power cannot be added for the reason already noted—that the nominal interest rate is affected by the rate of price change, and so already includes an allowance for it. For a real asset, the cost of holding a dollar has two parts: the loss (or gain) in purchasing power because of rising (or falling) prices, plus the real return sacrificed on the alternative asset.

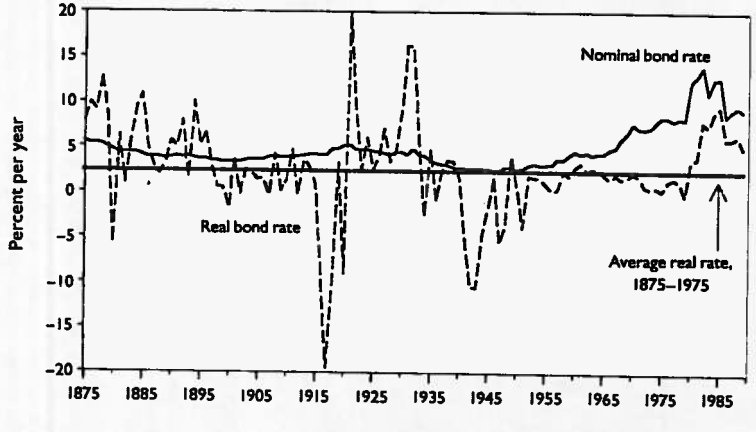
Over long periods of time, real returns on various classes of assets do tend toward equality. But at any one point in time, real returns may vary widely for different classes of assets. Moreover, what people who acquire assets *expect* them to yield—the *ex ante* yield—may differ widely from what the holders do in fact receive—the *ex post* yield.

Figure 1 is a striking example. It plots over more than a hundred years the observed nominal yield each year on a collection of long-term securities, the actual *ex post* real yield year by year, and the average real yield for the century 1875–1975. The nominal yield is fairly stable for most of the period, while the *ex post* real yield fluctuates all over the graph. Clearly, the assets were not purchased in anticipation of such widely variable yields. Such assets are typically held for long periods, so that the *ex post* yield for a holder of these assets was actually much less variable than the year-by-year calculations show. Indeed, the relatively stable nominal yield implies that the holders of the securities expected, *on the average,* zero inflation. And, until World War II, they were right: the price level in the United States in 1939 was roughly the same as in 1839.

The lines on the graph gradually change character after World War II, as the public came to recognize that inflation was more than a passing phenomenon. The nominal

Figure 1

Nominal and Real Bond Interest Rate, Annually, 1875–1989



yield rose to incorporate that recognition, and the real yield became more stable, started rising toward the long-term average, and then overshot it in the 1980s.

Given the wide range of assets that are alternatives to holding money, it is a great simplification to speak of *the* cost of holding money. There is in fact a vector of costs, depending on the particular alternative considered. And that is still an oversimplification. Even for a single asset, there is a range of possible yields, both nominal and real. Uncertainty about the nominal yield of alternative assets is one reason for holding money—there is little uncertainty about its nominal yield. But there remains uncertainty about the real yield of money. Empirically, high rates of inflation tend to be more unstable than low rates of inflation. As a result, both the level and the instability of inflation discourage the holding of money. That is why during periods of high and uncertain inflation, as in Chile in 1975, real cash balances are reduced so low, even though that greatly increases the costs of transactions.

Reconciling Supply and Demand

We have come a long way from our initial simple question: what determines how much we can buy with the green-backed five-dollar bill we started with? We are ready to return to that question by putting together the two blades of the monetary scissors, supply and demand.

In our hypothetical world in which paper money is the only medium of circulation, consider first a stationary situation in which the quantity of money has been constant for a long time, and so have other conditions. Individual members of the community are subject to enough uncertainty that they find cash balances useful to cope

with unanticipated discrepancies between receipts and expenditures. But these uncertainties average out, so that the community as a whole wishes to hold as cash balances an amount equal to one-tenth of a year's income.

Under those circumstances, it is clear that the price level is determined by how much money there is—how many pieces of paper of various denominations. If the quantity of money had settled at half the assumed level, every dollar price would be halved; at double the assumed level, every price would be doubled. Relative prices would be unchanged.

This very hypothetical and unreal situation, from which we shall shortly depart, brings out sharply one special feature of money: its usefulness to the community as a whole does not depend on how much money there is. For almost all goods and services, the utility derived from them depends on their physical quantity, on the number of units. For money, it does not. Doubling or halving the number of dollars simply means that the numbers written on price tags are doubled or halved. When gold ruled the monetary roost, there was much talk about whether there would be enough gold to serve as monetary reserves. That was the wrong question. In principle, one ounce would be enough. It would not physically circulate, as most gold did not, but claims to it could be issued in any fractional denominations—for example, one-billionth of an ounce—that were convenient.

The reason it was the wrong question is that no important or interesting issues of monetary theory arise in the hypothetical situation of constant demand and supply. As David Hume wrote more than two centuries ago, "it is of no manner of consequence, with regard to the domestic happiness of a state, whether money be in a greater or less

quantity" ([1742] 1804b, p. 305). As he also said, what matters is changes in the quantity of money and in the conditions of demand for money.

Let us suppose, then, that one day a helicopter flies over our hypothetical long-stationary community and drops additional money from the sky equal to the amount already in circulation—say, $2,000 per representative individual who earns $20,000 a year in income.* The money will, of course, be hastily collected by members of the community. Let us suppose further that everyone is convinced that the event is unique and will never be repeated.

To begin with, suppose further that each individual happens to pick up an amount of money equal to the amount he or she already holds, so that all find themselves with twice the cash balances they had before.

If everyone simply decided to hold on to the extra cash, nothing more would happen. Prices would remain what they were before, and individual incomes would remain at $20,000 per year. The community's cash balances would be 10.4 weeks' income instead of 5.2.

But people do not behave in that way. Nothing has occurred to make it more attractive for them to hold cash than it was before, given our assumption that everyone is convinced the helicopter miracle will not be repeated. (In the absence of that assumption, the appearance of the helicopter might increase the degree of uncertainty anticipated by the members of the community, which in turn might change the demand for real cash balances.)

Consider the representative individual who formerly held 5.2 weeks' income in cash and now holds 10.4 weeks'

*The helicopter example, and the rest of this section, is based on Friedman (1969, chap. 1).

income. He could have held 10.4 weeks' income before, if he had wanted to, by spending less than he received for a sufficiently long period. When he held 5.2 weeks' income in cash, he did not regard the gain from having $1.00 extra in cash balances as worth the sacrifice of consuming at the rate of $1.00 per year less for one year, or at the rate of 10 cents less per year for ten years. Why should he think it worth the sacrifice now, when he holds 10.4 weeks' income in cash? The assumption that he was in a position of stable equilibrium before means that he will now want to raise his consumption and reduce his cash balances until they are back at the former level. Only at that level is the sacrifice of consuming at a lower rate just balanced by the gain from holding correspondingly higher cash balances.

Note that the individual has two decisions to make:

1. To what level does he want to reduce his temporarily enlarged cash balances? Since the appearance of the helicopter did not change his real income or any other basic condition, we can answer unambiguously: to the former level.

2. How rapidly does he want to return to the former level? To this question we have no answer. The answer depends on characteristics of his preferences that are not reflected in the stationary equilibrium position.

We know only that each individual will seek to reduce his cash balances at some rate. He will do so by trying to spend more than he receives. However, one person's expenditure is another's receipt. The members of the community as a whole cannot spend more than the community as a whole receives. The sum of individual cash balances is equal to the amount of cash available to be held. Individuals as a whole cannot "spend" balances; they can only transfer them. One person can spend more than he

receives only by inducing another to receive more than he spends. They are, in effect, playing a game of musical chairs.

It is easy to see what the final position will be. People's attempts to spend more than they receive will be frustrated, but in the process these attempts will bid up the nominal value of goods and services. The additional pieces of paper do not alter the basic conditions of the community. They make no additional productive capacity available. They alter no tastes. They alter neither the apparent nor the actual rates at which consumers wish to substitute one commodity for another or at which producers can substitute one commodity for another in production. Hence, the final equilibrium will be a nominal income of $40,000 per representative individual instead of $20,000, with precisely the same flow of real goods and services as before.

It is much harder to say anything about the transition. To begin with, some producers may be slow to adjust their prices and may produce more for the market at the expense of nonmarket uses of resources. Others may try to make spending exceed receipts by taking a vacation from production for the market. Hence, measured income at initial nominal prices may either rise or fall during the transition. Similarly, some prices may adjust more rapidly than others, so that relative prices and quantities will be affected. There might be overshooting and, as a result, a cyclical adjustment pattern. In short, without a much more detailed specification of reaction patterns, we can predict little about the transition. It might vary all the way from an instantaneous adjustment, with all prices doubling overnight, to a long-drawn-out adjustment, with many ups and downs in prices and output for the market.

We can now drop the assumption that each individual happened to pick up an amount of cash equal to the amount he had to begin with. Let the amount each individual picks up be purely a matter of chance. This will introduce initial distribution effects. During the transition, some people will consume more, others less. But the ultimate position will be the same.

The existence of initial distributional effects has, however, one substantive implication: the transition can no longer, even as a conceptual possibility, be instantaneous, since it involves more than a mere bidding up of prices. Let prices double overnight. The result will still be a disequilibrium position. Those individuals who have picked up more than their pro rata share of cash will now have larger real balances than they want to maintain. They will want to spend the excess, but over a period of time, not immediately. On the other hand, those individuals who have picked up less than their pro rata share will have lower real balances than they want to maintain. But they cannot restore their cash balances instantaneously, since their stream of receipts flows at a finite time rate.

This analysis carries over immediately from a change in the nominal quantity of cash to a once-and-for-all change in preferences with respect to cash. Let individuals on the average decide to hold half as much cash, and the ultimate result will be a doubling of the price level, a nominal income of $40,000 a year with the initial $2,000 of cash.

This simple example embodies most of the basic principles of monetary theory:

1. The central role of the distinction between the *nominal* and the *real* quantity of money.

2. The equally crucial role of the distinction between

the alternatives open to the individual and to the community as a whole.

3. The importance of attempts, as summarized in the distinction between *ex ante* and *ex post*. At the moment when the additional cash has been picked up, desired spending exceeds anticipated receipts (*ex ante*, spending exceeds receipts). *Ex post*, the two must be equal. But the *attempt* of individuals to spend more than they receive, even though doomed to be frustrated, has the effect of raising total nominal expenditures (and receipts).

Let us now complicate our example by supposing that the dropping of money, instead of being a unique, miraculous event, becomes a continuous process, which, perhaps after a lag, is fully anticipated by everyone. Money rains down from heaven at a rate that produces a steady increase in the quantity of money, let us say of 10 percent a year.

Individuals could respond to this steady monetary downpour as they ultimately did to the once-and-for-all doubling of the quantity of money, namely, by keeping their real balances unchanged. If they did this, and responded instantaneously and without friction, all the real magnitudes would remain unchanged. Prices would behave in precisely the same manner as the nominal money stock, rising from their initial level at the rate of 10 percent a year.

Again, while people could behave that way, they would not. Before the helicopter arrives, our representative individual could spend all his income and add nothing to his cash balances, yet the cash balances would remain equal to 5.2 weeks' income. They remained constant in real as well as nominal terms because prices were stable. Storage costs and depreciation costs were zero, as it were.

Now that the representative individual is getting cash

from the helicopter, he can keep his *real* cash balances at 5.2 weeks' income from the sale of services only by adding all the extra cash to his *nominal* balances to offset rising prices. However, the money from heaven seems to be a bonanza enabling him to do better. If he reduces his cash balances by $1.00 over a year, he can now increase his consumption at the rate of $1.10 per year, whereas before he could have increased his consumption at the rate of only $1.00 a year. Since he was just on the margin before, he will now be over the margin. Storage and depreciation costs are now 10 cents per dollar per year, instead of zero, so he will try to hold a smaller real quantity of money. Suppose, to be specific, that when prices are rising at 10 percent a year, he wants to hold one-twelfth instead of one-tenth of a year's proceeds from the sale of services in cash balances, that is, 4⅓ weeks' income instead of 5.2.

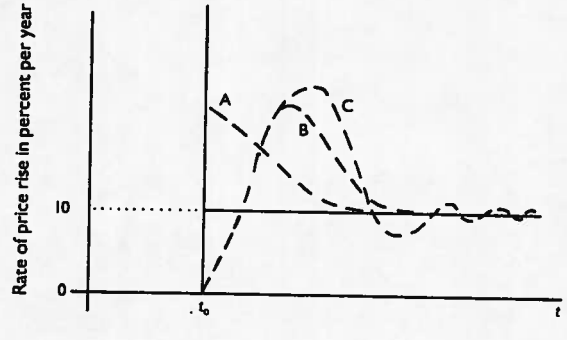
We are now back to our earlier problem. To each individual separately, it looks as if he can consume more by reducing his cash balances, but the community as a whole cannot do so. Once again, the helicopter has changed no real magnitude, added no real resources to the community, changed none of the physical opportunities available. The attempt of individuals to reduce their cash balances will simply mean a further bidding up of prices and income, so as to make the nominal stock of money equal to one-twelfth instead of one-tenth of a year's nominal income. Once that happens, prices will rise by 10 percent a year, in line with the increasing amount of money. Since both prices and nominal income will be rising at 10 percent a year, real income will be constant. Since the nominal quantity of money is also rising at 10 percent a year, it stays in a constant ratio to income—equal to 4⅓ weeks of income from the sale of services.

Attaining this path requires two kinds of price increase: (1) a once-and-for-all rise of 20 percent, to reduce real balances to the level desired when it costs 10 cents per dollar per year to hold cash; (2) an indefinitely continued inflation at the rate of 10 percent per year, to keep real balances constant at the new level.

Something definite can be said about the transition process this time. During the transition, the rate of inflation must average more than 10 percent. Hence, inflation must overshoot its long-term equilibrium level. It must display a cyclical reaction pattern. In Figure 2, the horizontal solid line is the ultimate equilibrium path of inflation. The three broken curves after t_0, the date at which the quantity of money starts to rise, illustrate alternative possible transitional paths: curve A shows a single overshooting and then a gradual return to the permanent position; curves B and C show an initial undershooting, then an overshooting, followed by either a gradual return (curve B) or a damped cyclical adjustment (curve C).

The necessity for overshooting in the rate of price change and in the rate of income change (though not

Figure 2



necessarily in the level of either prices or income) is, I believe, the key element in monetary theories of business cycles. In practice, the need to overshoot is reinforced by an initial undershooting (as in curves B and C). When the helicopter starts dropping money in a steady stream—or, more generally, when the quantity of money starts unexpectedly to rise more rapidly—it takes time for people to catch on to what is happening. Initially, they let actual balances exceed long-run desired balances, partly out of inertia; partly because they may take initial price rises as a harbinger of subsequent price declines, an anticipation that raises desired balances; and partly because the initial impact of increased money balances may be on output rather than on prices, which further raises desired balances. Then, as people catch on, prices must for a time rise even more rapidly, to undo an initial increase in real balances as well as to produce a long-run decline.

While this one feature of the transition is clear, little can be said about the details without much more precise specification of the reaction patterns of the members of the community and of the process by which they form their anticipations of price movements.

One final important detail. Implicitly, we have been treating the real flow of services as if it were the same on the final equilibrium path as it was initially. That is wrong, for two reasons. First, and less important for our purposes, there may be permanent distributional effects. Second, and more important, real cash balances are at least in part a factor of production. To take a trivial example, a retailer can economize on his average cash balances by hiring an errand boy to go to the bank on the corner to get change for large bills tendered by customers. When it costs the retailer 10 cents per dollar per year, rather than nothing, to

hold an extra dollar of cash, there will be a greater incentive to hire the errand boy, that is, to substitute other productive resources for cash. This will mean both a lower real flow of services from the given productive resources and a change in the structure of production, since different productive activities may differ in cash-intensity, just as they differ in labor- or land-intensity.

Our simple hypothetical helicopter example brings out clearly a phenomenon—some might call it a paradox—that is of the utmost importance in the actual course of events. To each individual separately, the money from the sky seems like a bonanza, a true windfall gain. Yet when the community has adjusted to it everyone is worse off in two respects: (1) the representative individual is poorer because he now has a reserve for emergencies equal to 4⅓ weeks' income rather than 5.2 weeks'; (2) he has a lower real income because productive resources have been substituted for cash balances, raising the price of consumption services relative to the price of productive services. This contrast between appearance to the individual and the reality for the community is the basic source of most monetary mischief.

The Famous Quantity Equation of Money

The preceding discussion can be summarized in a simple equation—an equation that was envisaged by scholars centuries ago, was stated carefully and precisely in the late nineteenth century by Simon Newcomb, a world-famous American astronomer who, on the side, was also a great economist, and was further developed and popularized by Irving Fisher, the greatest economist the United States has ever produced. In Fisher notation, the equation is

$$MV = PT .$$

M is the *nominal quantity of money.* As we have seen, it is currently determined in the United States by the Federal Reserve System. V is the *velocity of circulation,* the number of times each dollar is used on the average to make a purchase during a specified period of time. If we restrict purchases to final goods and services, and if the public holds 5.2 weeks of income in cash, as in our example, then the velocity is 10 times per year, since a year's income (equal to a year's purchases of final goods and services, which include savings) is ten times the quantity of money.* V is determined, as we have seen, by the public according to how *useful* it finds cash balances and how much it *costs* to hold them. The product of M and V, the left side of the equation, is total spending or income.

On the right side, P is an *average price,* or an index of the average price, of the goods and services purchased. T stands for *transactions,* to be interpreted as an index of the total quantity of goods and services purchased. Fisher, in his original version, used T to refer to all transactions—purchases of final goods and services (like bread purchased by the final consumer), intermediate transactions (flour purchased by the baker), and capital transactions (the purchase of a house or a share of stock). In current usage, the item has come to be interpreted as referring to purchases of final goods and services only, and the notation has been changed accordingly, T being replaced by y, as corresponding to real income.

*Again, I should warn that this is a reasonable number only for money defined as currency or base money. In recent years, base money velocity has been around 15 or 20, decidedly higher than earlier. For a definition like the current U.S M2, velocity is much lower, about 1.3 a year.

As written, the equation is an identity, a truism. Every purchase can be viewed in two ways: the amount of money spent and the quantity of a good or service purchased multiplied by the price paid. Entering the amount of money on the left side and quantity times price on the right, and adding up those sums for all purchases, we have a standard case of double-entry bookkeeping. As in double-entry bookkeeping in general, the truism is highly useful.

Consider once more our original question: What determines how much you can buy with the greenbacked five-dollar bill we started with? *Nothing can affect* P *except as it changes one or more of the other items in the equation.* Will a boom in the stock market, for example, change how much you can buy with a five-dollar bill? It will reduce the amount you can buy (raise P) only if it leads the Fed to create more money (increases M), or induces people to hold lower real cash balances, perhaps because they think the alternatives have become more attractive (raises V), or reduces the quantity of goods and services available for purchase, perhaps because workers are paying less attention to their work and more to the stock ticker (lowers T). The stock-market boom can raise the amount you can buy (lower P) only if it has the opposite effect, and clearly there are all sorts of possible combinations.

As this exceedingly simplified example illustrates, the equation is a useful way to organize an analysis of the effect of changes in circumstances. In short, Fisher's equation plays the same foundation-stone role in monetary theory that Einstein's $E = mc^2$ does in physics.

Changes in the Quantity of Money

In the real world, money does not drop from helicopters. When money consisted largely of physical commodities like gold and silver, new discoveries and technological advances were a major source of changes in the quantity of money. Chapter 3 discusses the effects of the nineteenth-century discoveries of gold and silver, the most dramatic of which were the Californian (1849) and Australian (1850s) discoveries; the opening of the Comstock Lode (1859), rich in silver and gold; and, later in the century, the Alaskan and South African finds. Chapter 5 discusses the effect on William Jennings Bryan's political career of the most dramatic technological change, the perfection of the cyanide process for extracting gold from low-grade ore.

Consider, in light of the helicopter fable, the effect of the flood of gold from California and Australia in the 1850s. Like those who were quickest to pick up the helicopter money, the first to extract the gold were clearly enriched. My favorite example comes from a visit to a major Australian gold-mining town, now preserved as a tourist attraction. One antique document on display was an advertisement for ice from Walden Pond. Ice, cut in the winter from Walden Pond in Massachusetts, was loaded in sawdust into the hulls of ships, which then sailed around the tip of South America and across the wide Pacific—about fifteen thousand miles—to Melbourne, where the ice was unloaded onto carts and dragged by horses some hundred and more miles to the gold-mining community, to satisfy the desire for cold drinks of the lucky and newly wealthy gold miners!

The gold from California and Australia, being first spent where it was found, attracted people and goods (like the ice) from all over the world by bidding up prices. As this occurred, the gold came to be distributed around the rest of the world and ended by raising prices in all gold-standard countries. As in the helicopter fable, it took a long time for the effects of the discoveries to work themselves out. As they did, the initial wide discrepancies in prices were reduced.

Also as in the fable, the effect on individuals was very different from the effect on the community at large. The lucky persons who first extracted the gold were clearly enriched. But what about the community at large? At the end of the process, the community was worse off. The appeal of the lottery involved in the several gold rushes meant that the resources spent to extract the gold from the earth, transport it to distant lands, mint it into coins, and bury it in bank vaults were almost surely greater in value than the new gold. Some of the new gold doubtless went into jewelry, gold plate, and the like. This part, at least, provided a continuing source of utility. But the rest of the gold, used as money, mostly meant only that prices were higher than they otherwise would have been. As David Hume wrote in 1742, "augmentation [in the quantity of money] has no other effect than to heighten the price of labour and commodities. . . . In the progress towards these changes, the augmentation may have some influence, by exciting industry, but after the prices are settled . . . it has no manner of influence" (1804a, p. 314). "Exciting industry" may have produced some increase in output, but it is hard to believe that any such increase could have offset more than a trifling part of the cost in resources of the additional money.

While the welfare effects of the gold discoveries were almost surely negative, *it does not follow that the existence*

of a gold standard—or, more generally, a commodity standard—is a mistake and harmful to society. True, such a standard does involve the cost of digging the gold out of the ground in one part of the world in order, in effect, to bury it in another. However, we have seen that having a widely accepted medium of exchange is of critical importance for any functioning complex society. No money can serve that function unless its nominal quantity is limited. For millennia, the only effective limit was provided by the link between money and a commodity. That link provided an anchor for the price level. Departures in general were, in Irving Fisher's words, "a curse to the country involved." As noted earlier and discussed in more detail in chapter 10, the world is now engaged in a great experiment to see whether it can fashion a different anchor, one that depends on government restraint rather than on the cost of acquiring a physical commodity. That experiment is less than twenty years old as I write—young even on a personal time scale, let alone on a historical time scale. The verdict is far from in on whether fiat money will involve a lower cost than commodity money (see Friedman 1987 and also 1986).

I turn now to the other major source of changes in the quantity of money throughout history, and since 1971 the only source—action by government. From time immemorial, government has played a major role in the monetary system. One element of that role has been to seek to monopolize the coining of money. The objective was partly to standardize the money. The sovereign's seal on a coined piece of metal was intended to certify its weight and fineness and thus enable such coins to be used in transactions by tale, or number, rather than by weight, thereby reducing the cost of transactions. Another objective was to

earn seignorage, the mint's charge for converting bullion into coins.

Payment by tale, or count, rather than by weight greatly facilitated commerce.* But it also encouraged such practices as clipping (shaving off tiny slivers from the sides or edges of coins) and sweating (shaking a bunch of coins together in a leather bag and collecting the dust that was knocked off), whereby a lighter coin could be passed on at its face value. Gresham's law (that "bad money drives out good" when there is a fixed rate of exchange between them) came into operation, and heavy, good coins were held for their metallic value, while light coins were passed on. The coins became lighter and lighter, and prices rose higher and higher. Then payment by weight would be resumed for large transactions, and pressure would develop for recoinage.

Sweating and clipping were effectively ended by the milling of coins (the process of making the serrations around the circumference that we have come to take for granted), first used in 1663, and followed in Britain by the Great Recoinage of 1696 to 1698, which produced a much more homogeneous coinage.

A more serious matter was the attempt by the sovereign to benefit from his monopoly of coinage. In this respect, the Greek and the Roman experiences offer an interesting contrast. Though Solon, on taking office in Athens in 594 B.C., instituted a partial debasement of the currency, for the next four centuries (until the absorption of Greece into the Roman Empire) the Athenian *drachma* had an almost constant silver content (67 grains of fine silver until Alexander, 65 grains thereafter). It became the standard

*Much of the rest of this section is from Friedman (1974).

coin of trade in Greece and in much of Asia and Europe as well, and even after the Roman conquest the *drachma* continued to be minted and widely used.

The Roman experience was very different. Not long after the introduction (in 269 B.C.), of a silver *denarius* patterned after the Greek *drachma,* the prior copper coinage (*aes* or *libra*) began to be debased; by the beginning of the empire, its weight had been reduced from one pound to half an ounce. The silver *denarius* and the gold *aureus* (introduced about 87 B.C.) suffered only minor debasement until the time of Nero (54 A.D.), when almost continuous tampering with the coinage began. The precious-metal content of the gold and silver coins were reduced, and the proportion of alloy was increased to three-fourths or more of the coin's weight. By the end of the three-century-long debasement, the *denarius,* once nearly pure silver, had degenerated to little more than a copper coin with a thin wash at first of silver and then of tin. As an aside, it took less than a century for U.S. dimes, quarters, and half-dollars to go through the same life cycle. We do make progress.

The debasement in Rome (as ever since) was a reflection of the state's inability or unwillingness to finance its expenditures through explicit taxes. But the debasement in turn worsened Rome's economic situation and undoubtedly contributed to the collapse of the empire.

Debasement was necessarily a slow process, involving repeated recoinages and ultimately limited by the real cost of the baser metal. The spread of paper money in the eighteenth and early nineteenth centuries enabled the process to be speeded up. The bulk of the money in use came to consist not of actual gold or silver but of fiduciary money—promises to pay specified amounts of gold or silver. These promises were initially issued by private individuals or companies in the form of bank notes or transferrable book entries that have come to be called deposits. But gradually the state assumed a greater role.

From fiduciary paper money promising to pay gold or silver, it is a short step to fiat paper money—notes that are issued on the fiat of the sovereign, are specified to be so many dollars or francs or yen, and are legal tender, but are not promises to pay something else. The first large-scale fiat issue in a Western country occurred in France in the early eighteenth century (though there are reports of paper money in China a millennium earlier). Later, the French Revolutionary government issued paper money in the form of *assignats* from 1789 to 1796. The American colonies and later the Continental Congress issued bills of credit that could be used in making payments. These early experiments gave fiat money a deservedly bad name. The money was overissued, and prices rose drastically until the money became worthless or was redeemed in metallic money (or promises to pay metallic money) at a small fraction of its initial value.

Subsequent issues of fiat money in the major countries during the nineteenth century were temporary departures from a metallic standard. In Britain, for example, payment in gold for outstanding bank notes was suspended during the Napoleonic Wars (1797–1816). As a result, gold coin and bullion became more expensive in terms of paper. Similarly, in the United States the convertibility of Union currency (greenbacks) into specie was suspended during the Civil War and not resumed until 1879. At the war's peak, in 1864, the price of a twenty-dollar gold coin reached more than $50 in greenbacks.

Changes in the Demand for Money

As pointed out earlier, changes in demand for money can have the same effect as changes in the quantity of money. In speaking about changes in demand, however, it is important to distinguish sharply between those that arise from changes in the usefulness of cash balances, such as the spread of monetization or the increasing range of financial instruments available, and those that arise from changes in the cost of cash balances, such as changes in nominal interest rates and in the rate of price changes. In economic jargon, we must distinguish between shifts in the demand curve and movements along a demand curve for cash balances.

This distinction is important because changes in usefulness tend to proceed slowly and gradually. Many changes in cost conditions also come slowly, but when these changes are sharp, especially in interest rates and the rate of price change, they are generally the result of events put in train by prior changes in the supply of money. One recent example for the United States is the sharp rise in the rate of inflation and in interest rates during the 1970s and the subsequent sharp fall during the 1980s.

The conclusion is that substantial changes in prices or nominal income are almost always the result of changes in the nominal supply of money, rarely the result of changes in demand for money. (Chapter 8 discusses at greater length the key case of inflation.)

Conclusion*

Monetary phenomena have been subject to extensive study over centuries. A summary of some broad empirical findings from that research may help to focus the discussion of this chapter.

*This conclusion is largely from Friedman (1987).

1. For both long and short periods there is a consistent though not precise relation between the rate of growth of the quantity of money and the rate of growth of nominal income. If the quantity of money grows rapidly, so will nominal income, and conversely. The relation is much closer for long than for short periods.

2. Over short periods, the relation between growth in money and growth in nominal income is often hard to see, partly because the relation is less close for short than for long periods, but mostly because it takes time for changes in monetary growth to affect income. And how long a time is itself variable. Today's income growth is not closely related to today's monetary growth; it depends on what has been happening to money in the past. What happens to money today affects what is going to happen to income in the future.

3. For most major Western countries, a change in the rate of monetary growth produces a change in the rate of growth of nominal income about six to nine months later. This is an average that does not hold in every individual case. Sometimes the delay is longer, sometimes shorter. In particular, the delay tends to be shorter under conditions of high and highly variable rates of monetary growth and of inflation.

4. In cyclical episodes, the response of nominal income, allowing for the time delay, is greater in amplitude than is the change in monetary growth.

5. The changed rate of growth of nominal income typically shows up first in output and hardly at all in prices. If the rate of monetary growth increases or decreases, the rate of growth of nominal income and also of physical output tends to increase or decrease about six to

nine months later, but the rate of price rise is affected very little.

6. The effect on prices, like that on income and output, is distributed over time, but it comes some twelve to eighteen months later, so that the total delay between a change in monetary growth and a change in the rate of inflation averages something like two years. That is why it is a long row to hoe to stop an inflation after it has been allowed to start. It cannot be stopped overnight.

7. Even after allowance for the delayed effect of monetary growth, the relation is far from perfect. There's many a slip over short periods 'twixt the monetary change and the income change.

8. In the short run, which may be as long as three to ten years, monetary changes affect primarily output. Over decades, on the other hand, the rate of monetary growth affects primarily prices. What happens to output depends on real factors: the enterprise, ingenuity, and industry of the people; the extent of thrift; the structure of industry and government; the relations among nations; and so on.

9. One major finding has to do with severe depressions. There is strong evidence that a monetary crisis involving a substantial decline in the quantity of money is a necessary and sufficient condition for a major depression. Fluctuations in monetary growth are also systematically related to minor ups and downs in the economy but do not play as dominant a role as other forces. As Anna Schwartz and I put it: "Changes in the money stock are . . . a consequence as well as an independent source of change in money income and prices, though, once they occur, they produce in their turn still further effects on income and prices. Mutual interaction, but with money rather clearly the senior partner in longer-run movements and in major cyclical movements, and more nearly an equal partner with money income and prices in short-run and milder movements—this is the generalization suggested by our evidence" (1963, p. 695).

10. A major unsettled issue is the short-run division of a change in nominal income between output and price. The division has varied widely over space and time, and there exists no satisfactory theory that isolates the factors responsible for the variability.

11. It follows from these propositions that *inflation is always and everywhere a monetary phenomenon* in the sense that it is and can be produced only by a more rapid increase in the quantity of money than in output. Many phenomena can produce temporary fluctuations in the rate of inflation, but they can have lasting effects only insofar as they affect the rate of monetary growth. However, there are many possible reasons for monetary growth, including gold discoveries, the financing of government spending, and the financing of private spending. Hence, these propositions are only the beginning of an answer to the causes and cures for inflation. The deeper question is why excessive monetary growth occurs (see chapter 8).

12. A change in monetary growth affects interest rates in one direction at first but in the opposite direction later on. More rapid monetary growth at first tends to lower interest rates. But later on, the resulting acceleration in spending and still later in inflation produces a rise in the demand for loans, which tends to raise interest rates. In addition, higher inflation widens the difference between real and nominal interest rates. As both lenders and borrowers come to anticipate inflation, lenders demand, and borrowers are willing to offer, higher nominal rates to offset the anticipated inflation. That is why interest rates

are highest in countries that *have had* the most rapid growth in the quantity of money and also in prices—countries like Brazil, Argentina, Chile, Israel, South Korea. In the opposite direction, a slower rate of monetary growth at first raises interest rates but later on, as it decelerates spending and inflation, lowers interest rates. That is why interest rates are lowest in countries that *have had* the slowest rate of growth in the quantity of money—countries like Switzerland, Germany, and Japan.

13. In the major Western countries, the link to gold and the resulting long-term predictability of the price level meant that, until sometime after World War II, interest rates behaved as if prices were expected to be stable and neither inflation nor deflation was anticipated. Nominal returns on nominal assets were relatively stable, while real returns were highly unstable, absorbing almost fully inflation and deflation (as displayed in Figure 1).

14. Beginning in the 1960s, and especially after the end of Bretton Woods in 1971, interest rates started to parallel rates of inflation. Nominal returns on nominal assets became more variable; real returns on nominal assets, less variable.

CHAPTER 3

*The Crime of 1873**

I am persuaded history will write it [the Act of 1873] down as the greatest legislative crime and the most stupendous conspiracy against the welfare of the people of the United States and of Europe which this or any other age has witnessed.

—SENATOR JOHN H. REAGAN (1890)

[The demonetization of silver] . . . was the crime of the nineteenth century.

—SENATOR WILLIAM M. STEWART (1889)

In 1873 we find a simple legal recognition of that [the demonetization of silver] which had been the immediate result of the act of 1853.

—JAMES LAURENCE LAUGHLIN (1886)

You shall not press down upon the brow of labor this

*I am indebted for helpful comments on earlier drafts to Michael D. Bordo, Conrad Braun, Phillip Cagan, Joe Cobb, Harold Hough, David Laidler, Hugh Rockoff, and, as always and especially, Anna J. Schwartz. In addition, David D. Friedman and an anonymous referee on behalf of the *Journal of Political Economy* made a number of very helpful suggestions for revision.