2 Survey of the Literature

2.1 Before Keynes

2.1.1 Representative money and neutrality

It is only recently, with the collapse of Bretton-Woods, that modern money has broken its last connections with the world of commodities. Throughout almost all its prior history money has been commodity money.\footnote{Precious metals have been the commodity of choice among civilized societies because of properties like durability, divisibility and low ratio of weight to value. But primitive societies have used a wide variety of commodities as money. The strangest perhaps is the stone money used by the Yap people of the western Pacific islands. Gillilland (1975).} It is only in the past few decades that the people of the world have started using a means of exchange and a unit of account that does not correspond to any object usable for consumption or production.

We might then expect naively that monetary theory developed for a time when money was commodity money would be of little use to us today. Thankfully, the truth is very different. There is much in the work of the old masters of monetary theory that is strictly true only in a world where money is intrinsically useless. In a way theirs was a theory ahead of its time. This theory owes its existence to two factors. First, the breaking away of money...
from the world of commodities has not been sudden. It is the culmination of the development of credit and central banking over a period of centuries. Second, even in a purely commodity money world, once exchange is monetised to the extent that the non-monetary uses of the money-commodity become minor compared to its monetary use it is no longer possible to treat that commodity at par with other commodities. In determining its value we must take into account 'monetary' factors. The stronger these two forces become the more fruitful it becomes to ignore the commodity aspect of money and treat it as an object *sui generis*.

A clear early articulation of this idea is to be found in David Burne’s two essays, 'Of Money' and 'Of Interest' (Burne, 1777/1985, Essays III and IV of Part II). Burne says:

"Money is not, properly speaking, one of the subjects of commerce; but only the instrument which men have agreed upon to facilitate the exchange of one commodity for another" (Burne, 1777/1985, page 282).

This was not literally true in Hume’s time. Gold and silver were certainly subjects of commerce for the miners who produced them and the jewellers who purchased them for non-monetary purposes. Any theory developed from Hume’s premise would have required qualification for these factors then. It does not do so now.

If money is merely a token, and is understood by everyone to be merely a token, then what matters are the transactions that these token allow to be made. As long as this set of possible transactions does not change, the de-
tails of how many and what forms of token are to be used to carry out these transactions should have no real consequences. In contemporary terminology, agents should not suffer from "money illusion".

"Money having chiefly a fictitious value, the greater or less plenty of it is of no consequence, if we consider a nation within itself; and the quantity of specie, when once fixed, though ever so large, has no other effect, than to oblige every one to tell out a greater number of those shining bits of metal, for clothes, furniture or equipage, without encreasing [sic] any one convenience of life [...] as these metals are considered chiefly as representations, there can no alteration arise, from their bulk or quantity, their weight or colour, either upon their real value or their interest." (Hume, 1777/1985, page 297)

The observation that the exact amount of money tokens should not have any real consequences is a counterfactual statement about different possible state of the world—one with more money and one with less. To Hume and, as we shall see, to many other after him this has suggested a causal statement about a given economy, namely that a change in the quantity of money leads to a proportionate change in prices and leaves all real variables unchanged. Once again in contemporary terminology, we would say that money should be neutral. If we add to this the belief that the quantity of money is an exogenous variable we have the Quantity Theory of Money which claims that the primary determinant of the price level is the quantity of money.
2 Survey of the Literature

A counterfactual proposition like the absence of "money illusion" cannot by itself be an explanation for a causal claim like neutrality or the quantity theory. What is needed is a chain of actions by agents and the consequences of those actions that link the increase in the quantity of money to the increase in the price level. The need for building such a causal chain was felt by all serious students of the Quantity Theory and motivated important works of money theory such as Wicksell's *Interest and Prices* and Keynes' *Treatise on Money*. While progress has been made on understanding the consequence of a change in money stock, as we shall see in our discussion on Keynes and latter economists, the difficulty of this task has often led economists to attempt a short-cut by appealing to the notion of equilibrium to cut out the dynamical process of adjustment. In the end this approach turns out to be inadequate to establish the neutrality proposition or the Quantity Theory. We illustrate this failure in a simple model of an exchange economy. Our framework can be seen as a simplification of that of Patinkin (1965).

2.1.2 A simple exchange economy

The model

Consider a competitive exchange economy with the following characteristics. There are two dates, 0 and 1. We model risk by assuming that there are $S$ states of nature on date 1 so that there are $S + 1$ date-state pairs in all. There are $G$ (physical) goods and $H$ households.

Our model differs from the basic general equilibrium model in two re-
2 Survey of the Literature

spects. Firstly, we do not require that there be a complete set of contingent futures markets. Rather we assume that there are \( J \) exogenously given kinds of assets, where \( J \) may be less than \( S \). We assume that asset returns are given in some unit of account and each asset is characterised by the exogenously specified amount \( y^t \) that it pays in each of the date-state pairs \( t = 1, \ldots, S \). One of these assets can outside money in which case its payoff is one unit of itself in each future state. Households are forced to transfer their purchasing power between the date-state pairs using only these assets as best as they can.

Our second generalisation is that of not assuming perfect foresight or any of its variants. Therefore much of our discussion remains valid regardless of the particular expectational assumptions chosen.

Because of the sequential nature of our economy, each household has to satisfy multiple budget constraints. They are,

\[
\sum_i p_i^{t,0} (x_i^{t,0} - \omega_i^{t,0}) + \sum_j q_j (z_j^t - \bar{z}_j) \leq 0 \quad (2.1)
\]

\[
\sum_i p_h^{i,s} (x_h^{i,s} - \omega_h^{i,s}) \leq \sum_j y_j^{i,s} z_h^j \quad \text{for } s = 1, \ldots, S \quad (2.2)
\]

where \( p_i^{t,0} \) and \( q_j \) are the prices on date 0 of good \( i \) and asset \( j \) respectively. For each household \( h \), \( x_h^{i,t} \) and \( \omega_h^{i,t} \) are the quantity consumed and the endowment respectively of good \( i \) in date-state \( t \), and \( p_h^{i,s} \) is the price that it expects for that good in date-state \( s > 0 \). \( z_h^j \) is its holding and \( \bar{z}_h^j \) its endowment of the asset \( j \). In what follows, we shall usually drop the subscript for the household where no confusion is likely to arise.
2 Survey of the Literature

While the economic interpretation of eq.(2.1) is clear enough, eq.(2.2) means that households do not plan to default. A more detailed analysis would replace this equation with some mechanism which penalises default. However, any such mechanism would have to distinguish between planned default and default which takes places because of incorrect expectations. Our equation can be taken to summarise a situation where this distinction can be made perfectly and wilful default attracts an infinite penalty.

Coming to equilibrium conditions, the first thing that we require is the equalisation of demands and supplies for both assets and goods on date 0, i.e.

$$\sum_{h}(x_{h}^{i,0} - \omega_{h}^{i,0}) = 0 \quad \text{for } i = 1, \ldots, G$$

$$\sum_{h}(z_{h}^{j} - z_{h}^{j}) = 0 \quad \text{for } j = 1, \ldots, Q$$

(2.3)

If our equilibrium concept is that of temporary equilibrium then this is all that we demand. However, another equilibrium concept that is often applied to such models is that of Radner (1972) which requires that households have common expectations and that the plans formed by different agents on the basis of these expectations be mutually consistent. This requires that apart from (2.3) we impose the additional conditions

$$\sum_{h}(x_{h}^{i,s} - \omega_{h}^{i,s}) = 0 \quad \text{for } s = 1, \ldots, S \text{ and } i = 1, \ldots, G$$

(2.4)

Note, however, that Radner equilibrium does not require that agents have identical beliefs about the probabilities of the different date-states occurring.
Absence of money illusion

Consider two economies which differ from each other only in the respect that if a household's endowment in the first economy is \((\omega_h, \xi_h)\), then its endowment in the second economy is \((\omega_h, \lambda \xi_h)\) for some \(\lambda > 0\). Now if \((x, z)\) satisfies (2.1) and (2.2) at prices \((p, q)\) in the first economy, then so does \((x, \lambda z)\) at prices \((\lambda p, q)\) in the second economy and vice-versa.

We posit that our households are free from "money illusion" by assuming that household's utility depends only on its consumption vector \(x\) and not on the portfolio it holds. It then follows that if \((x, z)\) is demanded at prices \((p, q)\) in the first economy then \((x, \lambda z)\) is demanded at prices \((\lambda p, q)\) in the second economy. This is so because the set of consumption vectors that each household can afford is the same in both situations and hence the best consumption plan must be the same. If portfolio \(z\) finances this consumption vector in the first economy then \(\lambda z\) finances it in the second economy. Therefore the excess supply of goods is the same in both situations while the excess demand of assets in the second economy is \(\lambda\) times that in the first economy. However, the aggregate endowments of assets in the second economy is also \(\lambda\) times the first. Therefore if \((p, q)\) is an equilibrium price vector for the first economy then \((\lambda p, q)\) is a equilibrium price vector in the second economy. The equilibrium allocation corresponding to this equilibrium is the same in both economies.
Neutrality of money

Our model is thus an economy of the sort Hume had imagined where money is purely representative. Yet the neutrality of money need not prevail in this model because:

- Our thought experiment requires asset endowments to be increased for all dates in the future. Just increasing the endowments on date $0$ is not enough.

- Only a proportionally greater money supply is not enough. The endowment of all nominal assets must increase in the same proportion. Thus, an increase in money supply through open market operations are ruled out.

- It is not enough for the aggregate supply of all assets to be higher proportionally. Each household’s endowment must increase in the same proportion.

- Equilibrium allocations will in general not be unique. In fact for almost all economies with incomplete markets there will be indeterminacy of equilibrium Geanakoplos and Polemarchakis (1986). Since we have provided no explanation of how the observed outcome may be chosen out of the multiplicity, a change in nominal endowments may modify observed outcomes even if it manages to leave the equilibrium set unchanged.
Conclusion

The representative character of contemporary money gives rise to the hope that we can understand some aspects of modern money by exploiting the consequences of this representativeness without needing to understand the details of how money functions. The belief in the neutrality of money and the Quantity Theory of Money is an expression of this hope. But we see that this hope cannot be fulfilled. Homogeneity of demand functions or symmetry of equilibrium sets cannot by themselves lead to any conclusion on the effects of policy interventions. The details of the working of money need to be understood after all.

2.1.3 Real effects of monetary changes

The need to understand the details of the monetary mechanism come not only from the theoretical problems of the Quantity Theory but also because a literal version of the Quantity Theory does not correspond to fact. If the only consequence of a monetary change were to be a change in the price level and if agents were free from "money illusion" then money would not matter at all and monetary economics would be a very unimportant subject. The interest in monetary questions arises precisely from the popular belief that money does affect real variables. To establish this statistically one faces the task of disentangling endogenous changes in money from exogenous ones and any method for doing so is liable to be questioned. Yet, many studies using very different methods and models, from Friedman and Schwartz
2 Survey of the Literature

(1971) to Christiano et al. (1999), find that money does affect real variables. How do we reconcile these facts with belief that money must be neutral? Here too Hume already had the crux of the idea that later economists have followed. Here is Hume facing up to the difficulty:

"Accordingly we find, that, in every kingdom, into which money begins to flow in greater abundance than formerly, every thing takes a new face: labour and industry gain life; the merchant becomes more enterprising, the manufacturer more diligent and skilful and even the farmer follows his plough with greater alacrity and attention. This is not easily accounted for, if we consider only the influence which a greater abundance of coin has in the kingdom itself, by heightening the price of commodities, and obliging every one to pay a greater number of these little yellow or white pieces for every thing he purchases." (Hume, 1777/1985, page 286)

Hume tries to work out the solution to this dilemma by imagining the consequences of an increase in the money holdings of some agents. In this he is being more ambitious than most modern writers on neutrality who deal with an equiproportionate increase in the money holdings of all agents. Neutrality is more plausible in this latter case than in Hume's since a change in the distribution of wealth which comes about as a result of some agents' money holding increasing while that of others don't might presum-

---

A notable exception is Archibald and Lipsey (1958) who seek to demonstrate the existence of a long-run equilibrium position that is independent of initial real balances.
ably change the equilibrium values of real variable. Nevertheless, Hume’s thought experiment anticipates both the strategy adopted in the modern approach and the problems it must face. Here is Hume’s solution, in a passage that is among the most controversial of his monetary writings (see Wennerlind (2005) for references to the discussion of this passage in the literature on the history of economic thought):

“To account, then, for this phenomenon, we must consider that though the high price of commodities be a necessary consequence of the encrease of gold and silver, yet it follows not immediately upon that encrease; but some time is required before money circulates through the whole state, and makes its effect be felt on all ranks of people. At first, no alteration is perceived; by degrees the price rises, first of one commodity, then of another; till the whole at last reaches a just proportion with the new quantity of specie which is in the kingdom. In my opinion, it is only in this interval or intermediate situation, between the acquisition of money and the rise of prices, that the encreasing quantity of gold and silver is favourable to industry. When any quantity of money is imported into a nation, it is not at first disposed into many hands; but is confined to the coffers of a few persons, who immediately seek to employ it to advantage. Here are a set of manufacturers or merchants, we shall suppose, who have received returns of gold and silver for goods which they sent to Cadiz. They are thereby enabled to employ more workmen than formerly,
who never dream of demanding higher wages, but are glad of employment from such good paymasters. If workmen become scarce, the manufacturer gives higher wages, but at first requires an encrease of labour; and this is willingly submitted to by the artisan, who can now eat and drink better, to compensate his additional toil and fatigue. *He carries his money to market, where he finds every thing at the same price as formerly, but returns with greater quantity of better kinds, for the use of his family.* The farmer and gardener, finding, that all their commodities are taken off, apply themselves with alacrity to the raising more; and at the same time can afford to take better and more cloths from their tradesmen, whose price is the same as formerly, and their industry only whetted by so much new gain. It is easy to trace the money in its progress through the whole commonwealth; where we shall find, that it must first quicken the diligence of every individual, before it encrease the price of labour.” (Hume, 1777/1985, page 286-287), emphasis ours.

Hume here reconciles the idea that “money matters” with his belief in the Quantity Theory by distinguishing a short-run adjustment process during which money has an influence on real variables from a long-run equilibrium position in which money is neutral. This immediately gives rise to a number of new problems.

First, why is it the case that an increase in the money stock of agents causes them to spend more on goods? If the agents had decided to hoard all
2 Survey of the Literature

the money then none of the subsequent steps in the argument would take place and money would be non-neutral. Thus we need to supply a theory of the demand for goods and assets in a monetary economy.

Second, even if an increase in money holding leads to an increase in the demand for goods, why is it the case that this increased demand leads to a change in real variables instead of working only to increase prices? The lynchpin of Hume’s argument is his claim that the workers who receive higher money wage “find every thing at the same price as formerly”. It is because some nominal variables are ‘sticky’ while others are flexible that nominal changes can lead to changes in real terms of trade and thus have real consequences. This stickiness of some nominal variables occurs in different forms in later theories where money has a real effect. Among the pre-Keynesians, (Fisher, 1922, Chapter IV) has a theory of a transition period in which money has a real effect because nominal interest rates are sticky. The short-period analysis in Kenes’ General Theory (Keynes (1936)) is of course the most famous example where nominal wages, prices and expectations are sticky.

An explanation of the non-neutrality of money in terms of stickiness does not help us identify the feature in economic behaviour or economic institutions where the Quantity Theory’s proposed path from the lack of money illusion to the neutrality of money breaks down. Therefore such an explanation cannot be fully satisfactory intellectually, though it may be empirically well-justified. We can join Lucas (1996) in asking Hume,

"Why, for example, does an early recipient of the new money
2 Survey of the Literature

'find every thing at the same price as formerly'? If everyone un­derstands that prices will ultimately increase in proportion to the increase in money, what force stops this from happening right away? Are people committed, perhaps even contractually, to continue to offer goods at the old prices for a time? If so, Hume does not mention it. Are sellers ignorant of the fact that money has increased and a general inflation is inevitable? But Hume claims that real consequences of money changes are "easy to trace" and "easily foreseen". If so, why do these consequences occur at all?"

The third problem that Hume must overcome is to ensure that his argu­ment for non-neutrality does not end up accomplishing too much. Once we have convincingly argued that a change in money can have real consequences, why should we continue to believe in the existence of a long-run position in which money would once again be neutral? This part of Hume's reasoning is the least developed at all and the most we can find in his works is a reiteration of the original idea that the lack of money illusion should lead to neutrality. But that is begging the question.

Progress on the last two problems that we have identified in Hume's argument had to await the Keynesian revolution. Between Hume's Essays and the General Theory contributions of lasting importance were made to the first problem of the demand of assets and goods and to the question of the exogenousity of money supply which separates the neutrality of money from the Quantity Theory proper. It is to these that we now turn.
2 Survey of the Literature

2.1.4 The demand for money

Hume began his analysis by assuming that an increment in money holding is not entirely hoarded and some fraction of it is spent on goods. This requires a theory of the demand of money and that of goods and the task of developing such a theory was taken up by later writers on the Quantity Theory. The most elaborate such theory that had developed before the General Theory was the cash-balance approach to the demand for money.

The antecedents of the cash-balance theory can be found in Irving Fisher’s The Purchasing Power of Money Fisher (1922). In this work Fisher introduced his famous ‘equation of exchange’ (Fisher, 1922, Chapter II, page 26)

\[ MV = \sum pQ \]

where \( M \) is the stock of money, \( V \) is its velocity of circulation, \( p \) and \( Q \) are the price and quantity respectively involved in a purchase transactions and the sum is taken over all such transactions in an economy in a given period.

As Fisher was well aware, by itself the equation of exchange is just a statistical truism. He, however, was to argue that an exogenous increase in \( M \) would not lead to a change in \( V \) and \( Q \) and would primarily manifest itself in a change in \( p \). For our present discussion it is his argument for the exogeniety of \( V \) which is of interest.

Fisher’s innovation was in seeing the velocity of circulation as arising out of the optimising behaviour of individual agents. He saw money as being held only for transaction purposes. An individual who had to decide on
2 Survey of the Literature

how much money to hold

"... will try to avoid carrying too little lest, on occasion, he be unduly embarrassed; and on the other hand to avoid encumbrance, waste of interest, and risk of robbery, he will avoid carrying too much." (Fisher, 1922, Chapter VIII, page 152)

Now imagine an individual who holds what he considers to be an optimal quantity of money. If the money stock held by this individual be increased, then the new stock would no longer be optimal—the individual would have much more money than he wants. A move to a new optimal point would involve holding a lesser quantity of money and, Fisher argues, the individual would make this adjustment by spending more on goods.³

The application of the logic of optimisation to the demand for money was taken much further by the Cambridge 'cash balance' school around Alfred Marshall (1929); Pigou (1917); Keynes (1923); Lavington (1921). These authors took the germ of the idea contained in Fisher and went on to present the determination of the value of money explicitly in a supply-demand framework. Moving away from the purely statistical equation of exchange, the Cambridge school focussed on relations that could be derived from the behaviour of agents. Thus, the central equation in Pigou (1917) is

\[ P = \frac{kR}{M} \]

³With the benefit of hindsight we know that all the additional expenditure need not be on freshly produced goods. It may as well be on second-hand goods or financial assets. But as long as the individual is not satiated in his consumption of goods, there must be some increase in the quantity of goods demanded as well. So the essentials of Fisher's argument are not affected.
where \( P \) is the Marshallian demand-price of money (the reciprocal of the price level of commodities), \( M \) is the stock of money held by an agent, \( R \) is an agent's real 'resources' and \( k \) is the fraction of these resource that the agent desires to hold as money.

In their discussion of the determinants of \( k \), the Cambridge school anticipated many of the factors considered by later monetary theory. Thus Pigou (1917) writes:

"This proportion \([k]\) depends on the convenience obtained and the risk avoided through the possession of such titles, by the loss of real income involved through the diversion to this use of resources that might have been devoted to the production of future commodities, and by the satisfaction that might be obtained by consuming resources immediately and not investing them at all"

Here we have the concept of an opportunity cost of holding money in terms of foregone return which was to play such an important role in Keynes' liquidity preference theory. We also have the idea that the decision of how much money to hold should be seen as being made simultaneously with the choice of an intertemporal consumption plan.

Another member of the Cambridge School, Lavington, discussed the effect on uncertainty on the demand for money in a way that foreshadows Keynes precautionary motive:

"Even in normal circumstances, therefore, the size of the stock of money held by a business man depends partly on the volume of his current transactions, partly on his individual business
outlook. The distinction is, perhaps, even more marked when this outlook changes. For when the general outlook is improving, rising confidence inevitably encourages business men to invest their resources more closely—to be less reluctant to "lock up their capital"; just as it encourages the banks to reduce the proportionate amount of their "cash" by expanding their loans. On the other hand, when confidence is declining, the amount of money which a business man requires to carry through his current transactions may be no greater than before; nevertheless his apprehension of the demands that may be made on him is likely to cause a great increase in his demand for money as a reserve against contingencies. If apprehension becomes general, there may follow a condition of crisis accompanied by a great and urgent demand for money, not so much to meet current payments as to provide against prospective payments whose volume is unknown." (Lavington, 1921, Chapter VI, page 32)

The Cambridge School’s insight of seeing the demand for money in choice-theoretic terms and thinking of this demand as being determined simultaneously with the choice of consumption and investment plans had the potential of leading to a theory that simultaneously explained monetary and real variables. However, this did not happen. The Cambridge school remained within the framework of the Quantity Theory, ignoring in practice the possible endogenous determinants of $k$ and treating it in effect as an institutionally determined quantity. It is not within the scope of the present
survey to fully investigate the historical reasons for this. But in terms of the structure of economic thought, the belief in the dichotomy between money and real variables certainly played a part in obstructing progress on the integration of monetary theory with the rest of economics. In essentials this dichotomy was just an elaboration of Hume's idea that since modern money was token money, the number of tokens in circulation cannot have any consequence over real variables. Compounding the conceptual difficulty was the implicit belief in Say's law and full employment, leading to the economists of that time believing that the value of real variables such as output and the interest rates were uniquely determined. If that was so then even if some of these variables entered into the demand for money, they could be taken as exogenous and hence ignored in the discussion of monetary questions.

Quantity theorists themselves realised that this picture was not fully consistent with evidence. The idea that "money did not matter" was even then hard to reconcile with the importance ascribed to monetary questions by the public and governments. Moreover, by the end of the 19th century the fluctuations in economic activity associated with the business cycle was well-known as was its association with fluctuations in money and credit. To explain these facts the classical dichotomy had to be broken somehow. But this was done in an ad hoc manner. In the passage quoted above, Hume had allowed for an increase in the stock of money leading to an increase in output because some prices do not change immediately. In Chapter IV of Fisher (1922) we find a theory of "transition periods" during which mon-
eterary changes have real effects due to the stickiness of nominal interest rates. There were many competing models of the "credit cycle". These development did not displace the classical dicothomy and Say's Law. They were still expected to hold in the 'long-run'. No well-articulated mechanism was presented to tie this long-run neutrality with short-run deviations from neutrality. As Samuelson (1968) writes in retrospect:

"To be sure, Hume, Fisher and Hawtrey had taught us that, under dynamic conditions, an increase in money might lead to 'money illusion' and might cause substantive changes—eg., a shift to debtor-entrepreneurs and away from creditors-rentiers, a forced-saving shift to investment and away from consumption, a lessening of unemployment, a rise in wholesale prices relative to sticky retail prices and wage rates, et cetera."

"But all this was at a second level of approximation, representing relatively transient aberrations. Moreover, this tended to be taught in applied courses in business cycles, money and finance, and economic history rather than in courses on pure theory. In a real sense there was a dichotomy in our minds; we were schizophrenics"

The Keynesian Revolution, which we discuss below, brought the analysis of the 'short-run', when money matters, into the centre of monetary economics. It did so on two grounds. First, at any moment of time we are in some short-run. The long-run can make sense only as a hypothetical limiting position of short-runs. The latter therefore has analytical priority
over the former. Second, even if such a long-run limiting position with classical properties exists, it would be irrelevant for practical purposes if the approach to it is too slow. The Keynesian Revolution was a watershed in monetary thought. The power and relevance of its methods overshadowed much earlier work on money. The exclusive analysis of a single short period in the General Theory however left it possible to argue that the Quantity Theory long-run could in fact be derived as a limiting position of Keynesian short-runs. If this were to be granted then the value of the Keynesian Revolution would be entirely dependent on how short in calendar time the 'short-run' was. Much of the counter-revolution against Keynes can be seen as an attempt to reinstate the Quantity Theory, but now in a way that can evade the charge of 'schizophrenia'.

2.1.5 Wicksell and his pure credit economy

Explaining the price level in terms of the quantity of money requires that the quantity of money be an exogenous variable. That this is so is not obvious in a modern monetary system where a large part of the means of exchange are not precious metal but banks deposits created by banks which are economic agents and whose decision on how much credit and deposits to create are naturally seen as endogenous. Proponents of the quantity theory have tried to get around this problem by either assuming that credit money bears a stable relationship to exogeneously determined non-credit money and/or that bank money should be regulated in a way that makes its quantity exogenous.
An entirely different approach was taken by Knut Wicksell in his model of a pure credit economy (Wicksell, 1907, 1936). Wicksell imagined an economy in which all payments are carried out by the transfer of bank deposits and there are no regulatory restrictions on the amount of deposits that can be created by the banks. Wicksell argued that in this world banks would be free to lend or borrow at any nominal rate of interest since they would be able to repay any debt by creating an equivalent deposit on their accounts.

However all possible nominal rates of interest are not equal. Suppose that the bank picks a nominal rate of interest that is smaller than the real return on investment in productive capital. Then there would be a large demand for loans by entrepreneurs willing to invest in productive capital and these entrepreneurs would bid up the price of productive capital. Though Wicksell does not make this explicit he is assuming that the expectations of future price levels are inelastic so that increase in the price of capital goods brings the nominal returns from them in line with the nominal rate of interest charged by the banks. However, in the next period price expectations would have been revised to correspond to the prices in the previous period and therefore prices would have to rise again to equate nominal returns on bank loans and nominal returns from investment in productive capital. This is Wicksell's famous cumulative inflationary process. A nominal rate of interest higher than the real return on productive capital would similarly set in motion a process of cumulative deflation.

Wicksell called the real rate of return on productive capital the 'natural rate of return'. Only when the nominal rate of return equals this natural
rate is the price level stable. Otherwise we have cumulative inflation or deflation.

Wicksell's pure credit economy made three major contributions to monetary theory. First, rather than seeing credit money as an encumbrance on a monetary system based on precious metals he had the foresight to see credit money as the dominant money form. He realised that in his time the dependence of the monetary system on precious metals would ultimately bring the cumulative process to an end and the primacy of Quantity Theory would be reasserted. But in our times Wicksell's theory comes into its own if we identify his bank with the central bank.

Wicksell's second contribution was in focusing on intertemporal terms of trade and the interest rate as determining the impact of money on the rest of the economy and breaking with the emphasis on the quantity of money. The picture of a banking system which fixes the nominal interest rate and lets the quantity of money adjust to demand corresponds much better to reality than assuming the quantity of money fixed.

The third major contribution that Wicksell made was in seeing the natural rate of interest not as a fixed quantity but as a quantity that depended on other economic circumstances. Thus an interest rate policy that wished to stabilise the price level could not be passive or mechanical but would have to respond to changes in economic circumstances. This in turn meant that a credit money system in which monetary policy was purely in human hands was much superior to a system where the quantity of money responded to external vagaries.
Wicksell's work had a major limitation too. He saw the effect of monetary policy as falling entirely on prices with quantities remaining unchanged. Keynes was to criticise Wicksell on the grounds that once one allowed deviations from full employment there is a natural rate of unemployment corresponding to each possible level of employment. A fuller understanding of modern money requires the integration of Keynesian unemployment equilibria with Wicksell's world of endogenous money.

2.2 Keynes

2.2.1 Treatise and chartalism

Keynes in his Treatise was to begin by pointing out that, "Money-of-Account, namely that in which Debts and Prices and General Purchasing Power are expressed, is the primary concept of a Theory of Money." Thus for him the unit of account function of money was primary. In discussing the origins of this unit of account, Keynes quoted with approval Knapp's idea of chartalism—that money is a creation of the state and it is the backing of the state that gives it its value. Since this belief is still persists in different forms, embodied in names like 'fiat money' for contemporary money, it is necessary to look at it in some detail.

What is this 'fiat' that the State issues in favour of fiat money? In all modern polities it consists of two components. First, currency printed by the government (and perhaps the notes issued by the central bank) is designated as legal tender, i.e. creditors are obliged to accept delivery of these
tokens as a discharge for debt. Second, the government demands taxes and carries out its other economic transactions in terms of this currency.

Let's begin with the legal tender property of government-issued currency first. By itself this property has no force. This is because the legal tender property of currency is relevant only with respect to those debts which have been contracted in terms of the unit of account corresponding to that currency. However, the private sector is not obliged to write contracts or publish price lists in terms of the government currency. As far as the laws are concerned, private agents may very well carry out its entire business in terms of a unit of account different from the official currency, thereby making the whole point of a legal tender moot. Therefore, unless prior economic reasons are given for why the private sector adopts the government-backed currency to specify its debts and prices, the legal tender property of currency can have no economic significance.

A simple example of this are the many Indian shops that sell imported goods like books. The price quoted by these shops is in dollars, though they accept payments in the rupee equivalent of the dollar price according to the prevailing exchange rate. The use of the rupee as a means of payments in this case is an artifact produced by laws restricting the development of a payments systems based on a foreign currency. If India had allowed its citizens to hold dollar-denominated bank accounts, as some non-US countries do, the buyer could well have paid for a book by writing a cheque for a dollar amount. There are instances in periods of hyperinflation when a much larger share of an economy has shifted over to denominated prices
and debts in terms of a foreign currency while the payments continue to be made in terms of the domestic currency.

In this context it is of no consequence that the dollar is specified as legal tender by the US government, since the transactions considered here happen outside the jurisdiction of the US government. In fact the place of the dollar can be taken by any other commodity as long as either there is a payments mechanism that works in terms of that commodity or as long as all parties can agree on what the equivalent of one unit of that commodity in terms of the means of payment is. In our example of the book shop this agreement is made possible by foreign exchange markets with well-defined prices which are publicly known, but any other organised commodity market would do as well.

It can be argued that neither of our examples is representative. The owner of an imported book shop faces a large part of his costs in terms of dollars and therefore hedges himself against exchange rate movements by quoting prices in dollars. His buyers acquiesce since for them books are a small fraction of their total expenditure. It is a form of risk sharing. Going further, it can be argued that outside of a hyperinflation most agents will find that most of their fixed obligations are in terms of the local currency and therefore the best way for them to avoid risk is to specify their other transactions in terms of the local currency as well.

But this putative counterargument suffers from a problem of circularity in its present form. Most agents have fixed obligations in terms of local currency only because most other agents have chosen to make claims in terms
of local currency. But that only tells us that one agent would not unilaterally want to shift to a currency different from that of its peers. But in an Indian economy where all prices and debts were denominated in dollars the force of preexisting obligations would now act in favour of new obligations being also contracted in dollars. The purpose of our examples above is to show that the lack of legal tender status for the dollar would not by itself be an obstacle to its use as a unit of account. So far it seems that the choice of a unit of account seems to be the same kind of choice as that of a language. It is advantageous to speak the same language as your neighbours, but the fact that a certain neighbourhood speaks Hindi rather than Japanese is purely a matter of accident.

Certainly the government could try to directly regulate the unit of account employed in private transactions. Doing so for a large enough, or an important enough, set of transactions would be sufficient to ensure that other transactions are also carried out in terms of the same unit of account. Preventing banks from creating foreign currency accounts would be one example of this. Note however that such economic regulation would go beyond merely providing legal tender status to the currency and would involve a more far-reaching intervention into the economic system. Such intervention is not universal.

There is however one way in which all governments intervene in the choice of the unit of account. This is through the choice of units in which they levy taxes, make purchases and hire workers. If we assume that the government can strictly enforce its tax demands then it would certainly en-
2 Survey of the Literature

sure the creation of a demand for the tokens in the form of which these
taxes are to be paid. However, this does not automatically establish these
tokens as the means of payment or their name as the unit of account within
the private sector. For it is still possible for the private sector to use some
other object both as its unit of account as well as its means of payment and
convert this object into the official currency at the last possible moment.
Since taxes and government purchases are only a small proportion of to­
tal private transactions, most agents would have no incentives to switch to
the government-backed unit of account if another unit of account is already
domínant in the private sector.

Thus a correct appreciation of the grain of truth that exists in chartalism
can be obtained only by recognising that the choice of a unit of account is
primarily driven by the pattern of mutual obligations and claims within the
private sector. The backing of the government can only be a stabilising force,
sustaining a unit of account already established in the private sector.

Keynes' *Treatise on Money* also contained major new insights into the port­
folio approach to the demand for money. We discuss them in the context of
their further development in his *General Theory*

2.2.2 General Theory

The *General Theory of Employment, Interest and Money* (Keynes, 1936) con­
tribution to monetary economics comes in two parts. The first is Keynes' theory of liquidity preference which gives an explanation for the demand for money and together with an assumption of an exogenously given sup-
ply of money determines the nominal rate of interest. This part of Keynes' theory, expounded primarily in Chapters 13–15 of the General Theory, is relatively well understood and its descendants are now found in elementary textbooks. The other part of Keynes' theory is concerned with how the nominal rate of interest and other nominal quantities can influence real variables like employment. The crux of Keynes explanation of this can be found in Chapter 17 of the General Theory which is notoriously obscure. We review both these parts of General Theory's contribution to monetary economics in turn.

2.2.3 Determination of the nominal interest rate

Demand and supply

Keynes' theory of liquidity preference is a theory of portfolio choice. He notes that households must make two kinds of choices. One, they must choose how much of current income to save rather than consume. Two, they must choose in what form to hold their total wealth. He faults classical economists for looking only at interest rates only as a determinant in the first, consumption-savings, choice and therefore seeing interest rates as being determined by the equality of demand and supply of loanable funds. But Keynes' argues that savings depend on incomes as well and once one leaves the classical world where the economy is always at full employment, the level of income itself becomes an endogenous variable. Therefore, by itself the equality of savings and investment cannot determine the rate of
interest. To complete the system, Keynes argues, we must examine the role
the rate of interest plays in the portfolio choice of households and firms.

Keynes' portfolio choice is a binary one: on one hand is money which
has the property of 'liquidity' and on the other hand there are all other as-
sets which Keynes clumps together under the category of bonds. What is
this 'liquidity' that money possesses? As far as Keynes' portfolio theory
is concerned 'liquidity' consists of three properties: money is the means of
payments in goods and asset markets, it has negligible carrying costs, and
its returns in terms of itself is certain. In Keynes' text the last is presented in
the form of the assumption that money does not pay any interest, but a de-
terministic positive rate of interest on money would not make a difference
to his argument. On the other hand the nominal return on bonds is taken to
be uncertain.

The first source of demand for money—the 'transactions motive' and a
part of 'precautionary motive'—arises from its role as the means of pay-
ment in goods market. Agents must hold a stock of money to pay for goods
and services, and greater the uncertainty regarding the timing and quantity
of payments the more money they have to hold for this purpose. For given
institutional arrangements, the amount of money needed for transactions
purposes would be an increasing function of transactions and hence an in-
creasing function of income. It would also be a decreasing function of the
nominal interest rate since the interest rate would be the opportunity cost
of idle stocks of money.

For Keynes, the transactions motive for holding money was less impor-
2 Survey of the Literature

tant compared to the desire to hold money as a store of wealth. Keynes apparently was thinking of asset markets with transactions costs since he breaks up the wealth demand for money into two parts. The first part, which is the remaining component of the 'precautionary' demand for money, arises from agents who wish to hold their assets in the liquid form of money in the anticipation of future investment opportunities rather than locking their wealth into presently available investment opportunities. This notion of 'liquidity' makes sense only if there are transactions costs involved in converting assets to money.

The lynchpin of Keynes' theory however is the second part of the asset demand, the 'speculative' demand for money. In a world where the nominal returns on money is zero, why would anyone hold money as an asset instead of holding assets with a positive expected nominal return? Keynes finds his solution in the uncertainty of the returns on bonds. The net returns to holding bonds consists of interest payments plus capital gains/losses. If the future price of bonds (or what amount to the same thing, the future rate of interest) is uncertain, then it is possible to contemplate the possibility of capital loss which would be larger than the interest payments and hence result in net negative returns.

Keynes works with a population of agents each of whom holds a point expectation regarding future bond prices. Those agents whose expectations of future bond prices is so low that they expect to incur a capital loss larger than the interest payments they receive prefer to hold money. These are the 'bears'. They are the source of the 'speculative' demand for money. The
other agents, whose expectation of future bond prices are high enough that they expect a positive net return from holding bonds, hold bonds. They are the 'bulls'.

How does the speculative demand for money behave as a function of the present nominal interest rate? Assume initially that agents' expectations of future bond prices are constant. Then an increase in interest rate would cause some agents who were expecting a negative net return earlier to now expect a positive net return since the higher interest rate is sufficient to cover the capital loss they are expecting. These agents switch from money-holding bears to bond-holding bulls. The demand for money goes down. Thus with constant expectations of future bond prices, the speculative demand for money will be a downward sloping curve.

In fact constant expectations are not necessary to derive a downward sloping demand curve for money. Inelastic expectations would do as well. Suppose the present price of a bond in \( P_t \) and the bond makes a coupon payment of \( a \) per period. Let the expected future price of the bond be a function \( P_{t+1}^e = \phi(P_t) \) of the current price. Then the net return from holding the bond will be:

\[
\rho = \frac{\phi(P_t) + a}{P_t} - 1
\]

Differentiating, we have

\[
\frac{d\rho}{dP_t} = \frac{P_t \phi'(P_t) - (\phi(P_t) + a)}{P_t^2}
\]

36
Thus the net returns on holding the bonds is a decreasing function of the current bond prices only if,

$$\frac{P_t \phi'(P_t)}{\phi(P_t)} < \frac{a}{\phi(P_t)}$$

(2.5)

The left-hand side of the above condition is the elasticity of an agent's expectation of future bond price with respect to changes in current bond prices. If the expectation function of all agents satisfy the bound in (2.5) then each agent would see then net return of bonds increase when the interest rate increases, i.e. the price of bonds falls. In that case we could expect agents on the margin to switch from money to bonds and we would have a downward sloping demand curve for money, just as in the case of zero-elastic expectations.

As we will see below, bounds on the elasticity of expectations like (2.5) play an important part in short-period models in ensuring that demand functions in terms of current prices behave as expected. As we shall see in detail in Chapter 3, such bounds are required to guarantee the existence of equilibrium in temporary equilibrium models. In economic terms, bounds on the elasticity of expectations are a reflection of agents' beliefs that there is a 'normal' range of prices so that beyond a point an increase or decrease in prices does not lead to an expectation of an increase or decrease of an even larger proportion in the future.

The transactions, precautionary and speculative motive for holding money together yield a downward sloping demand curve for money in terms of the nominal interest rate. Keynes assumed that the supply of money was
2 Survey of the Literature

exogenously determined by the central bank. The equilibrium between this demand and supply yield the nominal rate of interest.

Comparative statics

The effects of change of money supply can be traced using the usual comparative static methods. For example consider a given increase in money supply. The decrease in the rate of interest this causes would have to be just enough that the wealth of agents who switch from being bulls to bears is equal to the increase in money supply. From this Keynes argued that a diversity of opinion contributed to the effectiveness of monetary policy since he believed that a greater diversity of opinion would lead to a given change in the stock of money leading to larger changes in interest rates. But a little analysis shows that for this to be true ‘diversity’ needs to be defined in a special way.

Normalise the wealth of the community to be 1. Suppose as before that the current price of the bond is $P_t$ and it makes a coupon payment of $a$ next period. Assume for simplicity that each agent has zero-elastic point expectations of future price of the bond. Let $F(x)$ be the proportion of wealth held by those who believe that the bond price next period will be less than or equal to $x$. Then $F(x)$ is a distribution function in the probabilistic sense. Assume that it has a density $f(x)$. Let $M$ be the given money supply.

Since only those who expect the bond to give a negative net return hold
money the condition for equilibrium is:

\[ F(P_t - a) = M \]

Differentiating, we have

\[ f(P_t - a) \frac{dP_t}{dM} = 1 \]

or,

\[ \frac{dP_t}{dM} = \frac{1}{f(P_t - a)} \]

So, what Keynes means by a greater "diversity of opinion" is really a smaller density of opinion at the currently prevailing prices, which is a local concept. Keynes' proposition cannot be true if we understand diversity of the opinion in terms of some global property of the distribution \( F(\cdot) \) such as the variance. This is so because for any two pairs of distributions, the density of one cannot always be strictly less than the density of the other. Even if one distribution has a smaller variance than the other, for some values of \( P_t \) it will have a lower density as well. An example is easily furnished by plotting two normal distributions with different variance parameters.

**Extensions and criticisms**

Keynes' theory of the demand for money was put into a formal form by his successors. Baumol (1952) and Tobin (1956) framed the transactions demand for money in inventory-theoretic terms. Tobin (1958) reframed the speculative demand for money in terms of agents who hold probabilistic ex-
2 Survey of the Literature

expectations rather than point expectations about future bond prices and who therefore have both money and bonds in their optimum portfolio, rather than holding either only money or only bonds. Tobin’s paper laid the foundation of the modern theory of portfolio choice which was further developed by Markowitz (1952), Sharpe (1964) and others. These reformulations, however, did not change the essential economic nature of the Keynesian theory of the demand for money.

Looking at the Keynesian liquidity preference theory critically, we need to make two points. First, we need to ask what is ‘money’ for Keynes. For Keynes ‘money’ is something which has a deterministic return in the given unit of account in the short period. This is contrasted to ‘bonds’ whose return in the short period is uncertain. But this definition of money need not be coterminus with either the set of assets commonly accepted in exchange or the set of assets whose supply is controlled by the government. For example, there exist short-maturity treasury bills whose nominal return over a short period is certain, though they cannot be used as means of payment. There also exist short period commercial paper, trade credit etc. issued by the private sector which are neither issued by the government nor used as means of exchange. However, all these assets have certain short-period nominal returns. Then Keynes liquidity preference theory can give only a theory of the interest rate differential between these assets and assets of a longer maturity which may show capital gains or losses over the short period. But the Keynesian theory in the version given above can no longer explain the level of the interest rate itself. Second, the short-maturity fixed
nominal return assets we have discussed above dominate assets like currency and demand deposits which do not pay any interest at all. Therefore we need to fall back upon the use of these assets as the means of payment in order to explain the demand for them. Indeed, very shortly after Keynes work the beginnings of theory on these lines was done by Kalecki (1939, Chapter 5) and Kaldor (1939).

Second, Keynes assumes that the supply of money in his model is exogenous. This need not be true in a modern credit money economy. As Kaldor (1986), Moore (1988) and others have argued, it is much more reasonable to understand the behaviour of modern central banks as determining a rate of interest and then allowing the supply of money to respond to its demand. Kaldor argued that controlling the quantity of money would be inconsistent with the function of a modern central bank as a lender of last resort. Moreover, he argued that even if the central bank did try to control the quantity of money, the private sector would invent money-substitutes and alternative payment arrangements to augment the money supply. When initially formulated the endogenous money approach stood in contradiction to the dominant academic approach, deriving from the Quantity Theory, of treating the quantity of money as an exogeneously determined variable. However, of late economists have generally come to accept that it is the nominal rate of interest which is the policy variable (see for example Woodford (2003) which represents the current New Keynesian orthodoxy). This relegates the theory of liquidity preference to a sideshow. It is needed only to determine the quantity of money that is demanded at the rate of inter-
2 Survey of the Literature

The chosen rate of interest for the central bank. For understanding the level of activity and the consumption/production decisions of the private sector it is this rate of interest which must be considered.

2.2.4 Money and real variables

Whether we take the nominal rate of interest as being determined by the liquidity preference theory or as an exogenously given policy variable, we must still explain why a nominal quantity like the nominal rate of interest should influence consumption, investment or employment. By itself the liquidity preference theory of the rate of interest does not rule out a Wicksellian world where the nominal rate of interest determines only the rate of inflation while real variables like the level of employment and the real interest rate are determined by real factors. Keynes's own explanation of why this is not so, of why money matters, is to be found in Chapter 17 of the General Theory, the chapter devoted to the 'Essential Properties of Interest and Money'.

The starting point for Keynes is the observation that in equilibrium all assets in positive supply must earn the same rate of return. The numeraire in terms of which costs and payoffs are expressed does not influence this comparison. Suppose the return on two assets, with costs and payoffs measured in money terms is \( r_1 \) and \( r_2 \), then the return on the same two assets when the costs and payoffs are measured in wheat terms is \( r_1 - a \) and \( r_2 - a \) where \( a \) is the rate of increase of the money price of wheat. The second pair of returns would be equal if and only if the first pair are. In what follows we will
choose money as our numeraire, but this is purely a matter of convenience, not yet an essential property of money.

In his own exposition Keynes takes the stock of money as being determined by the central bank and then discusses how changes in this stock influence the nominal interest rate as well as how the nominal interest rate influences real variables. We believe that it will contribute to the simplicity of exposition, though possible at some cost of faithfulness to Keynes, to proceed in two steps. We begin by assuming that the nominal rate of interest is fixed by policy. Later we will consider the influence of the money stock on the nominal rate of interest.

**A given nominal rate of interest**

So to begin with we assume that the nominal rate of interest is fixed at some value \( i \). Assume that there is only one more asset in economy—a consumption-cum-capital good whose yield in terms of itself is a function \( r(q) \) of the quantity of the good produced. The supply price of the good is fixed in money terms because of given money wages; it is a function \( p_t(q) \) of the quantity produced. Agents have a point expectation \( p_{t+1}^e \) of the price of the capital good in the next period. Then equality of rates of return requires that:

\[
1 + i = [1 + r(q)] \left( \frac{p_{t+1}^e}{p_t(q)} \right)
\]  

(2.6)

With given \( i \) and \( p_{t+1}^e \) and appropriate forms for \( r(\cdot) \) and \( p_t(\cdot) \) eq. (2.6) determines the quantity \( q \) of the good produced. The nominal variable \( i \) has
determined the real quantity \( q \).

What is different here from the Wicksellian world? There are two crucial assumptions here that link up the nominal to the real world: first, the assumption of zero-elastic expectations regarding the future price of capital goods. Second, the assumption that the current price of the good is determined in money terms for a given level of output.

The first assumption, of zero-elastic expectations of future commodity prices, is not made explicit in Keynes' own treatment. However, as we have noted above while discussing the theory of liquidity preference, upper bounds on the elasticity of expectations are required to ensure that models of temporary equilibrium of the kind used by Keynes are well-behaved. By itself this does not lead to a conclusion different from Wicksell. If money were only a numeriare, then in a pure credit economy without outside money or prior nominal commitments, money wages and prices could move proportionately within a period without changing any real quantities. In that case in eq. (2.6) \( p_t \) would be an arbitrary quantity and not a function of \( q \). Then the equation could be solved for any \( q \) and any \( i \) and would serve only to determine the price level.

The crucial assumption in Keynes is then the second assumption, that the prices of commodities are given in money terms, that \( p_t \) is a determinate function of \( q \). It is this assumption that sets the Keynesian analysis apart from the quantity theory and the classical dichotomy. For Keynes himself, this pre-determination of prices in money terms is a result of stickiness of money wages. As Keynes notes, it is not enough that the wage bargain be
made in money terms rather than in real terms. What is required is that the money wage so negotiated should not be completely flexible.

Therefore to understand the theoretical contribution of Keynes, we need to understand why it is the case that money wages and prices are not completely flexible. We can distinguish between three different ways of approaching this problem.

The first and the simplest explanation of the stickiness of money wages and prices is an empirical one. Even a casual observation of the economic world shows that it is not populated by perfectly competitive firms and households whose plans are coordinated by auctioneers. Instead we have firms with market power and trade unions. Employers and employees enter into long-term contracts with payoffs specified in money terms. Firms do not change prices in response to every change in their economic environment. It is certainly a task for economics to understand why this is so, a task taken up seriously for example by macroeconomists of the New Keynesian school. But even if we fail to understand exactly why contracts are written in monetary terms, or why wages and prices are sticky, we would be justified on purely empirical grounds to choose a model with sticky money price and wages over a model where price and wages are completely flexible.

The second argument, found in Chapter 17 of the General Theory, is that an economy with flexible money wages and prices would not be stable. If the level of employment were to be determined by the exogeneously determined nominal rate of interest together with the propensity to consume and incentives to invest, as Keynes argues, then the level of employment in the
2 Survey of the Literature

economy would be at the full-employment level only by chance. Therefore if money wages and prices were perfectly flexible, so that they fell whenever employment was less than full employment and rose whenever employment was higher than full employment, money wages would be forever falling to zero or shooting to infinity unless the employment level happened to be at the full-employment level by sheer chance. The system would be violently unstable.

The crucial point in Keynes's argument above is that the level of employment does not change when wages and prices fall or rise. This has a number of caveats, some of which he himself recognised. First, a fall in money wages and prices leads to a fall in the transactions demand for money. If monetary policy consisted of holding the stock of money constant then this would lead to a fall in the rate of interest (unless we were in a state of absolute liquidity preference) and this would presumably lead to greater investment and greater employment. This counterargument does not work if monetary policy consists in holding the nominal interest rate fixed, as we have been assuming. Second, a fall in money wages and prices increases the real wealth of those who hold claims in denominated in nominal terms and this 'real balance effect' should increase consumption—a mechanism that was central to Patinkin's (Patinkin, 1965) attempt to reconcile Keynesian macroeconomics with neoclassical theory. However we should put counter to this the decreasing wealth and ultimately bankruptcy of those who have obligations denominated in nominal terms, which would act to reduce consumption. Finally, if expected future wages and prices do not fall
in the same proportion as current wages and prices, i.e. if expectations are inelastic, then the change in the intertemporal terms of trade should increase consumption and investment in the present period.

While the above points were part of lively debate on whether flexible wages and prices could ensure full employment, this debate is not germane to our present task of understanding why wages and prices are sticky. Suppose we grant Keynes' contention that an economy with flexible wages and prices would be extremely unstable. That only tells us that trying to 'reform' capitalism by making wages and prices more flexible would be a counterproductive quest. It would also tell us that if capitalism had been the result of conscious design, then the designer would have specified wages and prices to be inflexible. But capitalism is not a product of conscious design. It is a spontaneous system. Therefore the fact that flexible wages and prices would be fatal for the system as a whole cannot be a consideration influencing agents engaged in bargaining over wages and prices. There must be institutional factors in actually existing economies that cause employers and employees to agree on sticky money wage contracts in their own self-interest. These factors cannot be uncovered by counterfactual arguments about economies with flexible wages and prices.

There is a third strand in Keynes's argument in Chapter 17 of the General Theory which seems to suggest that it is the essential properties of money which cause the money wage to be sticky. Unfortunately this part of the argument is as obscure as it is tantalising. Keynes says,

What, then, would the position be if wages were expected to
be more sticky (i.e. more stable) in terms of some one or more commodities other than money, than in terms of money itself? Such an expectation requires, not only that the costs of the commodity in question are expected to be relatively constant in terms of the wage-unit for a greater or smaller scale of output both in the short and in the long period, but also that any surplus over the current demand at cost-price can be taken into stock without cost, i.e. that its liquidity-premium exceeds its carrying-costs (for, otherwise, since there is no hope of profit from a higher price, the carrying of a stock must necessarily involve a loss). If a commodity can be found to satisfy these conditions, then, assuredly, it might be set up as a rival to money. Thus it is not logically impossible that there should be a commodity in terms of which the value of output is expected to be more stable than in terms of money. But it does not seem probable that any such commodity exists.4

The best interpretation we can offer for this is on the same lines as Keynes earlier argument about infinitely rising or falling wages. Suppose that wages are sticky in terms of some commodity, say wheat, and employment is once again determined by the level of effective demand. Then either of two conditions must hold for the equilibrium to be determinate. The first is that stocks of wheat are not accumulated and the demand for wheat is determined by the level of aggregate employment. Then it must be the case that

4General Theory, Ch. 17 (IV)
profit maximising firms are willing to produce the demanded quantity of wheat at the going wheat wage. This is what Keynes meant by "cost of the commodity in question are expected to be relatively constant in terms of the wage-unit". Alternatively, the output of wheat is determined by profit maximisation at the given wheat wage and there are agents who hoard or unhoard the difference between the supply of wheat and its current demand. Such hoarding or unhoarding will need to continue as long as the level of demand for wheat, apparently taken by Keynes to be determined by the level of aggregate employment, does not match the profit-maximising supply. But such continuous hoarding can take place only if "its liquidity-premium exceeds its carrying-costs (for, otherwise, since there is no hope of profit from a higher price, the carrying of a stock must necessarily involve a loss)". Given his emphasis on the possibility of deficiency of effective demand, Keynes seems not to have considered the opposite case of continuous unhoarding since in that case the system would hit the lower bound of 0 regardless of the characteristics of the commodity in question.

It is possible to examine the details of this argument. In particular Keynes seems to have allowed no role for the relative price of wheat vis-a-vis other commodities in determining its demand. However, examination of such details is not very significant since Keynes' argument is subject to the same overarching criticism as his argument about perfect price flexibility. Behaviour of agents in a spontaneous system like capitalism cannot be explained by appealing to the consequences of that behaviour for the system as a whole. What a sticky wheat price would imply for the stability of the
system as a whole does not affect the decisions of an employer and an em­ployee bargaining over the terms of a wage contract. Only an examination of incentives at a micro level can provide an adequate explanation of the terms in which they decide to write the contract.

2.2.5 The stock of money and the interest rate

So far, we have carried out our discussion assuming that the nominal interest rate was exogeneously determined by the central bank. This allowed for a simplification of our analysis and also corresponds to the way monetary policy is carried out by most central banks today, where the policy instrument used is the nominal interest rate rather than the stock of money.

But what is it about money which allows the central bank to dictate the nominal interest rate to the private sector? According to Keynes it is the low elasticity of production and low elasticity of substitution of money.

By low elasticity of production, Keynes means that a fall in the value of money vis-a-vis other commodities cannot lead to the transfer of productive resources towards the production of money. In fiat money systems this seems to be a political fact: the government holds the monopoly of producing legal tender. However, the history of monetary institutions shows that things are not so simple. The development of paper and credit money was at least in part a response to an incentive to economise on the use of precious metals. And an increase in the velocity of circulation of money amounts to the same thing as an increase in its supply. To the extent that there is an opportunity cost for holding money, there will always be incentives for the
private sector to invent new ways to economise on the holding of money. Some economists such as Kaldor have considered this continuous innovation as an essential part of modern money. Therefore Keynes' assumption of a low elasticity of production of money must be seen as a part of the short-period framework of his analysis, where the speed of innovation in payments mechanisms is slow enough to be ignorable.

Even then, Keynes realises that low elasticity of production is a characteristic that money shares with all other non-producible goods. He believes that money stands apart from them because of its low elasticity of substitution. Since the demand for money depends on its value in exchange, an increase in the price of money in terms of other commodities increases the amount demanded for it. Thus, unlike ordinary commodities there is no tendency to substitute away from money as its relative price rises.

Here again there are subtleties. While the argument is certainly true in the case of the transactions demand for money, it needs more careful reasoning for the asset demand for money. Here the choice is between money and nominal bonds and the relative merits of the two is unaffected by changes in the value of money, so prima facie Keynes argument is borne out: as the value of money rises there is no substitution away from money. But both money and bonds are means for households and firms to transfer purchasing power between periods and the amount of purchasing power that they seek to transfer, i.e. the consumption-savings decisions depend on the intertemporal terms of trade. If agents hold unit elastic price expectations regarding the price level then these terms of trade do not change and the de-
mand for both bond and money increase in the same proportion as prices. But if expectations of the future price level are elastic, then a rise in the current price level, leads to the anticipation of an even greater rise in the future price level, shifting terms of trade against future consumption. This leads to an increase in current consumption hence leading to a decrease in the demand for both money and bonds. Since Keynes does not take this possibility into account, his argument would be strictly valid only when expectations of the future price level are inelastic or at worst unit elastic.

Keynes uses the inelasticity of production and substitution to argue that if the given money supply leads to a nominal rate of interest that is too high to ensure full employment, then the position cannot be rectified by changes in prices or other decisions by the private sector. Given the aggregate demand function, only an increase in the stock of money that reduces the nominal interest rate can increase the level of employment. And even this can fail if the economy is in a state of absolute liquidity preference.

### 2.3 Monetarism

During 1960s and the 1970s, Monetarism, represented by the work of Milton Friedman and his associates, as well as few other economists such as Brunner and Meltzer, was the major challenger to the neo-Keynesian Synthesis. Monetarist themselves did not define 'monetarism' and the term has been defined differently by different commentators. Many economists have defined it in terms of a very broad approach to monetary questions.
2 Survey of the Literature

For example, (Patnaik, 2008, page 27) says, "The defining characteristic of Monetarism is the proposition that the value of money in the short run is determined by the demand for and supply of it". Similarly Hahn (1982) says "Monetarism I take to be the doctrine that the perfectly competitive economy in Walrasian equilibrium is adequately descriptive of the world we live in...". This broad approach to defining monetarism may in part be motivated by the influence of monetarism on later developments in monetary economics and macroeconomics. Indeed, Tobin (1981) calls new classical macroeconomics "Monetarism II" in order to point out the continuity between the two schools of thought.

However, in order to better understand the role played by monetarism in the development of monetary economics, it may be useful to separate monetarism as it emerged in the Monetarist-vs.-Keynesian debates from its later intellectual descendants. Looking at the former, we are faced with the difficulty that monetarists, Friedman in particular, did not pinpoint exactly what separated them theoretically from the neoclassical synthesis. Through much of the debate the monetarists presented their criticisms in empirical terms. At the end of the debate, when Friedman finally presented a theoretical framework, it was not much different in appearance from the neoclassical synthesis (Friedman, 1970; Tobin, 1972). This has led some commentators to surmise that the Keynesian-Monetarist debate was at its root an empirical debate. This is not however the case. The monetarists indeed had a vision of the monetary system that was different from that of the Keynesians. However, this vision has to be inferred from the different contri-
2 Survey of the Literature

In one influential survey Laidler (1981) identified the following as the key characteristics of monetarism:

1. "A 'quantity theory' approach to macroeconomic analysis in two distinct senses (a) that used by Milton Friedman (1956) to describe a theory of the demand for money, and (b) the more traditional sense of a view that fluctuations in the quantity of money are the dominant cause of fluctuations in money income."

2. "The analysis of the division of money income fluctuations between the price levels and real income in terms of an expectations augmented Philips curve whose structure rules out an economically significant long-run inverse trade off between the variables."


4. "(a) Antipathy to activist stabilisation policy, either monetary or fiscal, and to wage and price controls, and (b) support for long-run monetary policy 'rules' or at least prestated 'targets', cast in terms of the behaviour of some monetary aggregates rather than the level of interest rates."

It is the first two characteristics that are the most relevant for the purpose of the present thesis. We look at them in turn.
2.3.1 Friedman's 'restatement' of the quantity theory

Among the influential early contributions by Friedman was his paper 'The Quantity Theory of Money: A Restatement' (Friedman, 1956). Friedman claimed that his paper was a continuation of an oral Chicago tradition in the quantity theory of money. However, the nature of his argument is very different from the traditional quantity theory investigations of the influence of changes in the quantity of money on the price level. Instead, Friedman presented his theory primarily as a theory of the demand for money.

Friedman sees the demand for money as arising from the portfolio choice of agents. In line with the earlier cash balance tradition, Friedman argues that agents hold a part of their wealth in the form of money primarily because money provides a service by facilitating transactions, though it may also yield a return in nominal terms, say in the form of interest on demand deposits. It competes for a share of the agents portfolio with other forms of holding wealth such as bonds, physical goods and human capital. Friedman writes the demand for money (assuming constant rates of return as

\[ M = f \left( P, r_b, r_e, \frac{1}{P} \frac{dP}{dt}, w; Y, u \right) \]

where \( P \) is the price level, \( r_b \) and \( r_e \) are the rate of return of bonds and equity respectively, \( w \) is the ratio of non-human to human wealth, \( Y \) is permanent income and \( u \) indicates factors that affect tastes and preferences.

So far this does not differ in spirit from Keynes's view of the demand for money as arising from the portfolio decision of agents, except for consider-
ing a larger set of assets. This similarity and an examination of the writing of earlier Chicago economists on the quantity theory led Patinkin to conclude (Patinkin, 1969) that Friedman’s paper was not so much a continuation of a Chicago quantity theory tradition as a continuation of Keynes’ monetary theory. As Patinkin says: “Friedman provided us in 1956 with a most elegant and sophisticated statement of modern Keynesian theory . . .”.

The theoretical part of Friedman’s analysis merely amounts to an assertion that the demand for money is a consequence of agents’ portfolio choice problem which can be studied in a utility maximisation framework. This however sidesteps the key issue in utility maximising models of money: the fact that the demand for money is not a primitive demand in the way the demand for wheat is. Rather, money is sought as an instrument for transferring purchasing power between dates and states and for carrying out transactions. The ‘utility’ of a stock of money is therefore an indirect utility and the dependence of this indirect utility on different economic variables would depend on the details of the economic environment. Since Friedman does not provide any of this detail, his casting of the portfolio choice problem in terms of utility maximisation brings no additional insight and does not constitute an advance over Keynes.

On the other hand, Friedman’s presentation lacks Keynes’ insights into the importance of uncertainty and the diversity of expectations in driving the demand for money. Keynes had carried out his analysis in a short-period framework, which though limited, had an advantage of being well defined in terms of involving only current variables and expectations of variables
one period ahead. Friedman appears to be working in terms of a long-term equilibrium since he uses 'permanent income' as the income variable in his model. But he does not provide to make explicit the definition of permanent income and the presumably 'permanent' price level and return variables. Neither is he explicit regarding his expectational assumptions. These omissions once again rob Friedman's exposition of any economic content.

In this paper Friedman distinguishes his work from the Keynesians not in terms of his theoretical framework, which he believes should be acceptable to "almost every economist", but on the following three claims:

1. The demand for money is highly stable.

2. There are factors affecting the supply of money which are independent of its demand.

3. There is no 'liquidity trap' and the demand for money is a function of a complex of a rate of returns and not just of a single interest rate.

Of these (1) is an empirical claim to which history has not been very kind. (2) and (3) are straw men—Keynesians, including Keynes himself, have taken the supply of money as being exogeneously given; the 'liquidity trap' was recognised by Keynes himself as a rare situation; the recognition of the variety of assets and rate of returns corresponding to them is only an extension of the portfolio approach to the demand for money. Further, neither of the three points made by Friedman is a consequence of his theoretical model.
2 Survey of the Literature

Even more importantly, Friedman's theory of the demand for money either by itself or with the assumption of an exogenous money supply cannot tell anything about how money influences the level of economic activity or the price level since more than one endogenous variable occurs in the demand function. However in the decade following Friedman (1956) Friedman did not follow up his demand function of money with clear cut specification of his theoretical framework for the rest of the economy. Instead, he based his policy recommendations on the empirical regularities he perceived in different economic time series. The resulting intense debate between the monetarists and the Keynesians engaged the macroeconomic profession and had major implications for public policy. In retrospect, these debates have not had a lasting impact on economic theory, mainly because of Friedman's reluctance to distinguish himself from Keynesians at the level of basic principles. Monetarism has had a more lasting impact on current economics through another of Friedman's interventions: his introduction of the concept of the natural rate of unemployment.

2.3.2 The natural rate of unemployment

Keynes' attention on a short period with fixed prices meant that the original Keynesian system did not have any account of price determination and thus could not be used to give an account of the price level. Over time, in the new Keynesian synthesis this gap was filled by the Phillips curve—a downward sloping relationship between the rate of change of money wages on one hand and unemployment rate on the other that took its name from
2 Survey of the Literature

Phillips study of the relationship between these two variables in the United Kingdom (Phillips, 1958). The most influential initial attempt to give a theoretical derivation of the Phillips curve, that by Lipsey (1960), saw it as arising from the slow adjustment of money wages to excess demand and supply in the labour market.

While Phillips' original curve posited a relationship between the rate of change of unemployment, the term Phillips curve was soon to be used more commonly for an analogous downward-sloping relationship between the rate of change of prices and the unemployment rate. If this curve could be assumed to represent a structural relationship, then aggregate demand management policies, including monetary policy, would be able to peg the unemployment rate at any desired level as long as the policymaker was willing to tolerate a higher steady rate of inflation. It was as such a menu of choices facing policy-makers that Keynesians such as Samuelson and Solow (1960) were to see the Phillips curve.5

It was this conception of the Phillips curve that Friedman was to question in his Presidential Address to the American Economic Association (Friedman, 1968)—an article that has been called "very likely the most influential article ever published in an economics journal" (Tobin, 1995). Friedman argued that monetary policy could not be used to permanently peg the unemployment rate at any arbitrary level. Indeed, there was only one level of the unemployment rate which could be so sustained on a permanent basis.

5Samuelson and Solow were prescient in noting both that the Philips curve might shift due to changes in expectations and that a period of high unemployment may raise the level of structural unemployment and thereby shift the curve. However these possibilities were not elaborated upon.
In analogy with Wicksell's "natural rate of interest", Friedman called it the natural rate of unemployment. He defined it thus:

"At any moment of time, there is some level of unemployment which has the property that it is consistent with equilibrium in the structure of real wage rates. At that level of unemployment, real wage rates are tending on the average to rise at a "normal" secular rate, i.e. at a rate that can be indefinitely maintained so long as capital formation, technological improvements, etc. remain on their long-run trends. A lower level of unemployment is an indication that there is an excess demand for labor that will produce upward pressure on real wage rates. A higher level of unemployment is an indication that there is an excess supply of labor that will produce downward pressure on real wage rates. The "natural rate of unemployment," in other words, is the level that would be ground out by the Walrasian system of general equilibrium equations, provided there is imbedded in them the actual structural characteristics of the labor and commodity markets, including market imperfections, stochastic variability in demands and supplies, the cost of gathering information about job vacancies and labor availabilities, the costs of mobility, and so on." (Friedman (1968), emphasis in the original).

Hahn and others were to criticize Friedman for speaking of a "Walrasian system" that embedded market imperfections, uncertainty and transactions costs. They pointed out that no way had been found to accommodate these
elements within the Walrasian system except in special cases. So, to speak of the "natural rate of unemployment" in terms of the equilibrium of such a Walrasian system would be meaningless.

However, a sympathetic reading of Friedman can see him as challenging the existing interpretation of the Phillips curve not on the basis of any particular theory of the economic equilibrium but from the point of view of the neutrality of money. In a world of agents free from money illusion, real variables such as the rate of unemployment should not be influenced by monetary variables. Therefore, in the quotation above Friedman asks us to replace the Philips relation between the rate of change of money wages and unemployment with an analogous relationship between rate of change of real wages and unemployment.

However, Fridman, the champion of "money matters" could not claim that money would always be neutral and that the rate of unemployment would always equal the natural rate. Instead, he argued that an attempt to achieve a lower level of unemployment that the natural rate would lead to accelerating inflation while an attempt to achieve a higher level of unemployment than the natural rate would lead to a decelerating rate of inflation.

How does this come about? Here is Friedman's argument:

"To begin with, much or most of the rise in income will take the form of an increase in output and employment rather than in prices. People have been expecting prices to be stable, and prices and wages have been set for some time in the future on

---

6This drastic neutralist step was to be taken by the real business cycle school later
that basis. It takes time for people to adjust to a new state of demand. Produces will tend to react to the initial expansion in aggregate demand by increasing output, employees by working longer hours and the unemployed, by takin jobs now offered at former nominal wages. This much is pretty standard doctrine.”

“But it describes only the initial effects. Because selling prices of products typically respond to an unanticipated rise in nominal demand faster than prices of factors of producing, real wages received have gone down—though real wages anticipated by employees went up, since employees implicitly evaluated the wages offered at the earlier price level. Indeed, the simultaneous fall \textit{ex post} in real wages to employers and rise \textit{ex ante} in real wages to employees is what enables employment to increase. But the decline \textit{ex post} in real wages will soon come to affect anticipations. Employees will start to reckon on rising prices of the things they buy and to demand higher nominal wages for the future. “Market” unemployment is below the “natural” level. There is an excess demand for labour so real wages will tend to rise towards their initial level.”Friedman (1968)

Thus changes in monetary policy can have an effect in the short-run, though the effect is at the expense of accelerating or decelerating inflation. Two things need to be noted about Friedman’s explanation. First, his is an equilibrium explanation: employers employ as many workers as they optimally wish to employ at what they perceive as the going real wage,
workers work as many hours as they wish at what they see to be the going real wage. This is different from Keynes' notion of unemployment equilibria with workers off the labour supply curve. Despite this, in Friedman the "market" unemployment rate can differ from the "natural" unemployment rate because of a limited form of money illusion on the part of the workers: they appear not to know the current price level and they use past price levels in order to translate a given money wage into their real wage. So, unanticipated inflation can fool workers can cause them to supply more labour. But workers catch up to the change in the inflation rate and therefore price have to rise faster to fool them again. To keep the unemployment rate constant the inflation rate has to accelerate or decelerate.

Thus, for Friedman, the original Philips curve is a special case that would be applicable when the inflation expected by the workers is zero. When this is not the case the Phillips curve must be the inflation-augmented Phillips curve, with unemployment deviating from its natural rate only in the presence of unexpected inflation or deflation.

Here the unemployment rate as a function of the inflation rate and not vice-versa. This is a consequence of Friedman's equilibrium approach to the labour market. The Phillips curve is no longer seen as a disequilibrium adjustment mechanism. It is now seen as the equilibrium supply curve for labour, with misperception of the inflation rate causing supply to deviate from its "natural" level. This view of a misperception based supply-curve of labour was combined with the assumption of rational expectations in Lucas (1972) to launch the New Classical 'counterrevolution' against Keynes-
2 Survey of the Literature

sianism. While the evidence (for eg. Barro (1978)) has not been kind to the New Classical supply curve, the legacy of Friedman's Presidential Address lives on in the form of a belief in the inability of monetary policy to affect real variables in the long run.

Friedman's contribution is important for reminding us of the tension that remains between the assumption of the lack of money illusion and the belief in the real consequences of monetary changes. Friedman's own resolution of this contribution does not go beyond the kind of arguments that had already been advanced by Hume and Fisher for example.