THE SIMPLE THEORY OF EFFECTIVE DEMAND

By Edward Nell*

Keynes set out to show that orthodox economic theory was flawed. To that end he accepted its postulates and general approach as the starting point for his argument. In particular he accepted marginal productivity theory as an account of the demand for labor, even though his own work in the Treatise had already suggested a different theory of distribution. Accepting marginal productivity theory committed Keynes to more than is generally realized, and marks a decisive difference between this model and Kalecki's. It opened the way to the IS-LM interpretation and to the "neo-Classical synthesis." Kalecki's model, by contrast, fits well with mark-up pricing and a realistic notion of market behavior, in both small and large scale industries.

The crucial commitment in the Keynesian approach comes at the very outset. The economic system is pictured as a circular flow of transactions between households and firms. Productive services are sold by households to firms and, with the proceeds, households in their turn purchase goods and services from the firms. Income is paid in exchange for productive services, which are combined to produce goods and services of equivalent value, which are sold back to the households as the income is spent. Obviously, as the flow is circular and continuous, it need be measured only at one point to establish its level for the entire circuit. Thus income is normally expressed as a sum of expenditures, necessarily equal to business receipts, which in turn exactly cover equilibrium factor payments.

But according to an older tradition of economics, and one that Kalecki explicitly drew on in his formulation of the theory of effective demand, income is not paid to factors for productive services. Instead, it is the distribution to social classes of claims to the surplus generated by production, where such claims are expressed in exchange value. The payment of income is not a market process of the same kind as the expenditure of income and its causes and consequences require a separate analysis. The significance of this change comes out very clearly in quite simple models, and at the risk of losing readers who are accustomed to more

*Graduate Faculty, New School for Social Research, New York, New York.

Warm thanks to Rosemary Rinder for detailed and constructive suggestions which saved me from many mistakes and greatly improved the presentation of the argument.
sophisticated fare. I would like to explore some very elementary problems in macroeconomics from the general standpoint developed by Kalecki and Joan Robinson.

/Income/

One of the striking features of the orthodox "income-expenditure model" is that it contains no expression at all for income. Output and expenditure are clearly indicated. Output is the product generated by employed labor working with the given stock of fixed plant and equipment, assumed constant in the short run. Output is the sum total of value added. Expenditure is the total for consumer spending, government spending, net exports, and new capital formation, minus transfers and inventory adjustments. Both these notions are represented in orthodox models, usually by the same symbol, Y, which is often called "income," but which it is not.

Aggregate income is the sum of payments received for productive services and, in the case of "unearned income," as a consequence of property rights. So it will be found by aggregating wages and salaries and the income of unincorporated business, on the one hand, and rents, interest, dividends and undistributed corporate profits, on the other hand. Making suitable adjustments for transfers and for defects in statistics, the result should be two broad categories: income from employment and income from property, each paid out regularly though not at the same rate per unit time as each other. Wages will normally be paid daily or weekly, salaries monthly, and profits quarterly or even yearly. Thus, we should be able to define two variables: W for employment income and P for property income; for simplicity we think of them as "wages" and "profits." The total income, \( Y^* \), will equal \( W + P \), and the income so paid out should be equal to the value of output. In general, the types of income aggregated into W will be paid out more frequently and will be adjusted quickly to the rate of sales, as these forms of income represent variable costs to firms. Property income, on the other hand, consists partly of fixed costs, such as debt servicing, and partly of residual income. Fixed costs must be paid at the contractually stated time, anywhere from quarterly to annually, while residual income can only be known for certain upon completion of an audit. So the incomes aggregated into P will not all vary so readily with sales, and none of them will be actually paid out in full during the short period.

- In short, wage income is available to finance money expenditures and is varied by business with the level of sales. But property income — profits — does not vary so readily, and is not paid out and may very well not even

---

1 No one else should be implicated in the specifics which follow, especially the diagrams, for which I (and many years of New School students) must take the responsibility. But the general approach is that of the post-Keynesian tradition. See Nell and Hollis [12], Introduction and Chapter 9. But the purposes of this paper are strictly limited. The determinants of investment, and particularly the role of finance, are not considered. In orthodox terms, this paper examines the "IS" side of macroeconomics.
be known during the short run, hence for the purposes of short-run analysis, property income is withheld from the flow of respending. This is not the same as calling it “retained earnings” or “business saving,” as these terms are normally understood. For in their usual sense, these terms refer to the final disposition of income. But regardless of that disposition, property income is not available to underwrite or influence spending in the short run.

The point is that wage income and other variable cost spending (purchases by the consumer goods sector of materials and replacements) is directly and immediately connected to spending for consumption goods (and by the consumption goods industries for materials, etc.). So when wage income varies, consumer spending varies pari passu. Working class families are assumed to have no other means of financing consumer spending, and are sufficiently needy that they will spend the whole of their income at once. By contrast, capitalist families with income from shares and lands are able, in the short run, to maintain their consumption spending since they possess assets which they can put up as security for loans. Short-run variations in property income will not affect capitalist consumption spending, because of this, and to keep the argument simple, I shall largely ignore capitalist consumption spending in what follows.\(^2\)

These points on consumption and wages are worth summarizing, together with plausible qualifications:

Consumption is strongly and stably related to wage income. It includes the consumption of wage earners, which amounts on average to the whole of their wages, though this will normally be arranged over the short run to minimize downward adjustments in consumption. It includes also the consumption of capitalists and quasi-capitalists, which probably has some constant rate of growth, not influenced by short-term changes in anything. It can, therefore, practically be disregarded in an analysis of short-run changes in employment, output, and income, although it should be remembered that capitalist consumption has a stabilizing influence on aggregate C, which in general provides a slight cushioning effect against changes in demand for consumption goods caused by a change in W.

The important point is that workers’ consumption expenditures — the bulk of total consumption demand — are cut (in practice, of course, as

\(^2\)To deal with it properly would require separating salaries from wages, on the one hand, and a closer look at consumer finance on the other. Both are beyond the scope of this paper.

\(^3\)The large and growing category of salaries in modern capitalism creates an empirical problem for this highly simplified theory of consumption. Some salaries (bank tellers, etc.) clearly fit into the theoretical concept of wages — they are variable and directly influence consumption spending; others (bank presidents, etc.) clearly contain large elements of profits, are quite stable, and do not influence spending. The empirical problems in separating categories, however, need not detract from the theoretical validity of the categories.
little as possible), when choices of capitalists create cutbacks in employment and income. Thus, consumption responds passively; it is dependent on the spending decisions of capitalists.

Three clear and distinct ideas: income, output and expenditures are often and confusingly represented by a single symbol in orthodox models. The relationships between the three are close but complex, and it will take some care to disentangle them. Most orthodox models clearly separate output from expenditure, even when both are indicated by the same symbol. But they do not represent income at all, which leads to problems in two directions. First, the costs and profits generated by producing output cannot be known until the correct income payments are determined. Second, the pattern of expenditure cannot be settled until the income of the various spending units is determined. So determining income and its distribution is essential in determining economically significant aspects of both output and expenditure. Let us explore these relations more closely.

In the aggregate, real income must equal the value of real output. This is a consequence of the property system. Everything produced must be owned, and the owner of the product — including all positive or negative liabilities incurred during production — is the firm. Hence, the residual earnings of the firm ensure that total income generated will equal the value of the total output produced. This does not, of course, ensure that money payments made during the period of production will exactly add up to the value of the product, let alone to the revenue generated by sales during that period. So there are several relationships here to keep clear:

\[ \begin{align*}
(1) & \quad \text{money claims} = \text{value of output} \\
(2) & \quad \text{money payments} = \text{value of output} \\
(3) & \quad \text{money payments} = \text{revenue from sales} \\
(4) & \quad \text{money claims} = \text{revenue from sales} \\
(5) & \quad \text{value of output} = \text{revenue from sales}.
\end{align*} \]

Only (1) is necessarily an equality. The total money value of claims against the product necessarily adds up to the value of the product. That is, the costs incurred by the firm for materials, labor, energy, depreciation, etc. are subtracted from the value of the output to give the firms net expected operating profit, and this residual, together with the costs, then constitutes the total of money claims issued, necessarily equal to the value of the product. There is a moral here for price theory, too. The price of the product will normally be set so as to generate a net operating profit of a certain size, so that it will stand to the invested capital of the firm in the normal ratio.

Now consider (2). The total money payments issued by the firm will include operating expenses, plus dividends distributed, plus debt servicing; and to make the case as strong as possible, add also the actual corporate retained earnings from the period. But this need not equal the value of the
product for the simple reason that everything produced during the period may not have been sold. (According to neo-Classical theory, this should lead to a fall in price, which would increase sales and reduce the value of output. But this would also lead to a revaluation of the entire output, requiring recalculation of the money payments.) Of course, in equilibrium the money payments, including corporate retained earnings, will equal the value of output: but even out of equilibrium when an audit has taken place and inventory value adjustments made, the money value of claims must equal the value of output. But money payments actually made during the period, including the banking of retained earnings, need not exactly equal the value of output.

On the other hand, money payments including retained earnings will tend to equal current revenue from sales. (3). For retained earnings, positive or negative, will make up the difference between costs paid out and current revenue. But this is not necessary in the same sense as (1), for costs and dividends could be pegged to the level of current production, determined by expected sales. So if current sales revenue were unexpectedly low, money payments could exceed revenue. Of course, this would require borrowing or running down cash balances, an important qualification. For money payments plus capital adjustments necessarily equal current revenue from sales; adding capital adjustments turns (3) into an accounting identity.

Finally, money claims will equal revenue from sales if and only if all production is sold for the planned price. (Since money claims necessarily equal the value of output, this covers: (5), value of output = revenue from sales.)

So we can say: in equilibrium, these will all hold. Out of equilibrium, however, only (1) and the modified version of (3) hold as accounting identities. Both (1) and (3) are consequences of the property system: (1) depends on the fact that everything produced is owned; (3) depends on the fact that all funds used must have a source. Payments cannot be made unless they can be financed: funds must come from somewhere and if not from sales revenue, then from capital or from borrowing. Thus (1) and (3) illustrate fundamental truths about the economic system.

These truths, however, sit awkwardly with the neo-Classical doctrine that income arises as a payment for a productive service. If that were true, all income payments would be costs of production and there would be no room for residual income. But it is precisely the claims to residual income that ensure the truth of (1). In the case of (3), on the one hand, revenue not equal to money payments requires an adjustment of assets, building up or running down capital holdings, to provide a source for the money payments; and on the other hand, revenue arises from spending which in turn must come from a source, either money payments of wages or profits, borrowing, or running down assets. Either way, revenue not equal to money payments requires a change in assets and liabilities, an adjustment.
of financial holdings, quite apart from any question of productive services or marginal products. The point is that the relation of income payments to the value of output and to the level of expenditure reflects the nature of the property system, the crucial characteristic of which is that one class owns the means of production, giving it title to the entire results of production, while the other class must work on a contractual basis for wages. The net residual claims of the propertied class consist, therefore, of the entire product Y, minus the contractual payments to employees, minus payments for materials and other inputs.

Output and Employment

Now let us examine a simplified model of production in such a class-based, or exploitative system, where output is constrained in the short run by industrial capacity. The obvious questions then are: first, how much of the industrial capacity will be used, and secondly, what will be the pattern of income distribution and employment? The questions are easy, but the answers may be harder to understand. In neo-Classical theory we expect to find equilibrium, the resting place of the system or the conclusion of a dynamic process, at the point where the market clears. And if the market does not clear, it will be because imperfections in it have prevented prices from adjusting sufficiently.

But we have just seen that the payment of income is not analogous to the buying and selling of final commodities. Both the payments to labor and those to capital have special features, intrinsic to the capitalist system, which differentiate them from each other and from market transactions for other goods. These features largely center on the employer-employee contract and entail that the system may settle into a stable position, as the result of a dynamic adjustment process, in which there is unemployment and excess capacity yet no prices have a tendency to move. nor need there be any “market imperfections” of the usual sort.

To understand this, the institutional setting must be clear. To see what is implied, assume a drastically simplified capitalist industrial society. There are two sectors: industry which produces capital goods both for itself and for agriculture, where factory-farms produce consumer goods. Capitalists own all the enterprise in both sectors and they or their loyal lackeys run them. Capitalist firms own the results of production, which they sell. In order to produce, they hire workers to operate their plant and equipment, paying them wages, which constitute the workers’ only source of income, and which workers spend in total on consumer goods.

The wage is at least sufficient to support workers and their families for a given period of time. (If it were insufficient, in the long run the working class would die off or emigrate.) In return, workers do the jobs their employers set for them during the time for which they have contracted. The faster they work, the more raw materials they process into finished goods, and the greater will be the employer’s potential profits. But the
faster they work the harder the job. Workers will, therefore, resist speed-ups and job reorganization. To maintain work discipline and keep productivity high, employers design equipment to run at the highest speed consistent with safety and accuracy in work, paying workers the minimum premium above subsistence necessary to induce them to accept the resulting working conditions. When sales fall off in the short run the plant cannot be redesigned, so parts will be shut down or run part-time, and workers on various shifts will be laid off until business picks up.

In a capacity constrained system, employment reflects the degree to which capacity is utilized. Demand in excess of capacity cannot be met, nor will it lead to the employment of more labor. Employment as well as output is constrained by capacity. No more men can be employed than there are places on the assembly line; jobs reflect the equipment installed in factories. Jobs will be offered as demand increases up to capacity; workers will be laid off as demand falls below capacity. Output will increase or decrease to meet demand as employment is varied. So there will be a definite functional relationship between employment and output, assuming that the level of output is uniquely correlated with the level of employment, that is, with the degree of capacity utilization, regardless for example of whether utilization reached that level by increasing or decreasing. When employment increases output will increase.

But how fast? Will output always increase in the same proportion when employment increases? Or will output increase more or less than in proportion? There are three possibilities: constant, increasing, and diminishing returns to the utilization of capacity.

But remember, there is no substitution here. Techniques of production, job definitions, organizational structures, plant and equipment, and business location are all fixed. What varies is the degree to which the given production system is utilized. The correspondence between the utilization of plant and equipment and the amount of employment depends on the rigidity of job definitions. The precise tasks are specified, together with the sequencing and the pace of work on the assembly line, as well as use of the labor force to clean up the shop, make repairs, and so on.

Utilization functions can be defined for each plant and aggregated to show the correspondence between output and employment for the whole society. What is the shape of the resulting function? Tradition would have it that successive doses of additional labor applied to fixed equipment will bring declining increments of output. But this belief is based on

---

*One can easily think of reasons why the direction of movement might have an effect. For example, when employment is diminishing, worker “morale” is likely to be low; that is, workers are likely to resist layoffs by slowdowns and work-to-rule. When employment is increasing, on the other hand, morale and productivity may rise together.

5 Aggregation, as always, causes problems. If the sectors have different capital-labor ratios, different levels of employment may be associated with a given level of aggregate demand, according to how it is made up of consumption and investment. This is a good reason for keeping the sectors separate, as set forth in the last section of the paper.
a misunderstanding. The argument for diminishing marginal returns depends on successive doses of employment being applied to efficiently utilized equipment. Each time labor was added, the use and perhaps also the nature of the equipment changed; only its amount as capital remained constant. Different levels of employment might be engaged in very different types of labor. Each level of employment, therefore, represented the "best practice" technique for that amount of labor applied to the given amount of capital, which, however, will generally be embodied in very different concrete forms at different points in the function. That such a function cannot be constructed has been demonstrated in the recent "capital theory" controversies; but this in no way affects the very different concept of a utilization function. However, for the same reason, the traditional arguments for diminishing returns have no weight either. Why should returns diminish when the plant is being run more nearly at the rate for which it was designed? Surely one might expect returns to be higher, user cost per unit lower, at the rate of output for which the engineers planned. And this is what some evidence seems to suggest.

But the evidence is not conclusive. The appearance of diminishing or increasing returns could result from different patterns of labor hoarding by an industrial system which was actually characterized by constant returns. Suppose that near the full employment point industry tended to hoard labor, while at lower levels of employment it became progressively more willing to vary the labor force in proportion to demand. The result would be a curve apparently exhibiting "increasing returns," but in fact lying along its whole length below the true constant returns line. (Figure 1) In the same way a ready willingness to dismiss workers, at high levels of employment coupled with labor-hoarding as employment falls towards the core of long-term experienced employees, will give rise to an impression of diminishing returns, the curve lying below and to the right of the true constant returns line. The two cases could be combined: labor hoarding could exist near full employment and again at low levels, where experienced long-term workers were threatened. This would yield a plausible function showing first diminishing and then increasing returns, just the reverse of the relationship assumed by conventional theory.

A conventional short-term production function shows alternative positions of fully adjusted equilibrium. Capital is embodied in the best choice technique, labor is organized efficiently, and so on. Hence, the system cannot move from one level of employment to another, for to do so would require reorganizing and perhaps rebuilding the capital stock. By contrast, a utilization function shows alternative levels of employment when a given industrial system is run more or less intensively, according to capitalist principles. So there is no difficulty describing movement from one level of activity to another. It is precisely the appropriate concept for analyzing short-run changes in activity.
The argument so far can be summed up concisely as follows:

1. \( Y = f(N) \), output varies with the level of employment of labor and capacity. Returns to capacity utilization are probably increasing, perhaps constant, unlikely to be decreasing.

2. \( Y = E \), employers adjust output and employment to the point where all output is sold at current prices.

3. \( Y = C - I \), definition of sectoral output.

4. \( E = W + P \), definition of class-based income streams.

5. \( W = C \), a theory of consumption.

6. \( P = I \), deducted from 3, 4, and 5.

It is the capitalists who initiate spending, who set investment and so determine the level of output and employment which generates the expenditures which assure that the output will be sold at prices the capitalists set. Observed profits will only roughly equal investment.
spending over rather long periods of time, especially if the consumption function is modified in the direction of greater realism, and because of the volatility of observed profits, which is greater than that of investment spending. But, the fact that the tautological result, \( P = I \), is only approximate, and only holds on average, does not weaken the theoretical point, which is that profits are passive, a result of the active investment decisions which essentially drive the whole economy.

Finally, output consists of sectoral outputs, just as incomes are the incomes of separate classes. We will return to this point later.

**Effective Demand**

The basic principles of effective demand can be illustrated most easily by concentrating on the constant returns case. First, the functional correspondence between employment and output needs to be explored further. Below a certain level, one would expect output to turn negative, indicating that workers use up more in the value of materials, energy, and user cost than they produce. However, this section of the utilization function can be ignored so long as the wage is always positive. For workers will not take employment in a capitalist system unless they are paid wages, which, having no assets to speak of, they must spend at once to support their families. Hence employment generates positive spending on consumption goods which in turn generates further employment. So long as the wage line lies above the output line, employment at such levels would generate demand in excess of output. (Figure 2) Conversely, the costs of such employment would exceed the returns from it. Either way, the range of values below \( N_o \) is infeasible. So \( N_o \) becomes the effective origin.

![Figure 2](image)

The slope of the wage line is shallower than that of the output line, reflecting the fact that the working force in the factory system as a whole produces more than is needed to sustain and reproduce itself, given the life style of the working class. Thus the ratio between the angles \( \beta \) and \( \alpha \),
The real wage bill here is represented by a straight line, indicating that the real wage rate is constant in the short run for different possible levels of employment. A change in the real wage would thus be shown as a change in the angle \( \beta \). Yet surely, according to traditional thinking, if employment is below capacity, the unemployed workers will drive the wage down? But this argument is defective. There is first a practical point: In the competitive areas of the economy business tends to be small scale, with less division of labor and lower mechanization. So the production process tends to depend on skilled and experienced workers and their crews, making it difficult to break into the cycle and fire presently employed workers to replace them by the unemployed at lower money wages. The loss in production and delay during training makes it inefficient. Unemployment will have an effect but not within the short period. In large scale industry, workers are more easily replaced since the division of labor, mechanization and automation have gone further, but so has unionization. In the short run such workers are protected by their union contract. Only in advanced industries with a highly developed division of labor and no unions, the least likely situation to be found in practice, would the emergence of short-run unemployment have a direct and immediate effect on money wages.

There is also a theoretical reason why the emergence of unemployment does not drive money wages down in the short run even when conditions are broadly competitive in a realistic sense. Workers do not set wage rates — employers do. Firms define jobs and announce openings at publicly stated rates of pay, which workers can accept or reject. If a firm cuts its money wage scales below the current market rate it will lose all or part of its labor force, unless all others cut by the same amount or more. But in the absence of collusion, no firm can know that others will also cut. A firm knows only that its sales have fallen off: it does not know, reliably, and certainly not at once, how its rivals are doing. And in the case of labor, it has to know not only what its rivals are doing, it has to know how all other markets are doing since workers can take employment in other industries as well. Thus, any given firm will be inhibited in offering lower money wages by uncertainty — the fear that others will not follow suit.\(^6\)

In any case, it is very easy to show changes in the real wage: the slope (angle) of the wage line changes. It is important to remember that this may happen as a result either of changes in money wages or changes in

\[\text{cut price} \Rightarrow \Delta \text{markup} \]

\[\text{Changes in } \Delta w = f\left(\Delta w, \Delta p, \Delta q\right)\]

---

\(^6\)This also acts to inhibit price cutting when sales fall off in the short run. If a firm is the first to cut price, it will attract sales, but by the same token it will then need its full labor force, which it will risk losing by announcing a wage below the prevailing rate. To cut price it must, therefore, be prepared to cut its markup, which as hypothesis has been set to earn some target rate of return over the expected economic lifetime of the capital equipment.
consumption goods prices. In either case the change is likely to be the result of interaction between a number of factors, and so should be derived precisely, rather than assumed. Hence, for the basic analysis, it is best to begin on the assumption of a given real wage, reflecting current prices and the established money wage.

Since capacity, output, and capacity employment are fixed, each can be taken as 100 percent, and the axes of the diagram (Figure 3) showing the functional correlation between employment and output can be measured in percentages. The constant returns diagram will then be a 45 degree line, and the wage, equal to worker consumption spending, will be given by a ray from the origin lying beneath the 45 degree line.

![Diagram](image)

**Figure 3**

Investment spending can now be shown. Investment decisions are taken by firms in the light of their long-run prospects — for new markets, developing new products, in the anticipation of growth of demand in existing markets, moving their present products into new regions or classes of the population, or in the case of capital goods, persuading new categories of business to adopt improved or new capital equipment. But current investment spending depends upon how rapidly businesses feel they should implement, carry through or complete the projects they have decided upon. It is a matter of timing; they have decided what to do, now the question is how fast to complete it. This will depend partly upon their ability to finance investment spending, and partly upon their confidence that the time is now ripe to bring these projects into operation. For our purposes, however, it will be best to leave the determinants of investment spending to one side, so as to concentrate attention on its effects. Assume for simplicity then, that investment spending is determined independently of the current level of employment and output. Aggregate demand will then be the sum of investment spending plus consumption spending.
(neglecting government spending and consumption by the capitalist class at this time). The intersection of aggregate demand and the employment utilization then gives the short-run equilibrium output and employment as percentages of capacity output and employment. Equilibrium output is \( Y^* \), and \( W^* \) is the corresponding total wage bill. The difference, \( Y^* - W^* \), then, is profit, \( P \); so the division of the vertical line rising from \( N \) to the 45 degree line gives the distribution of income between wages and profits, and also between consumption and investment. Notice that this diagram cannot deal with changes in productivity. Henceforth, output and employment will be measured in natural units. Other things being equal, then, the lower is the productivity of labor, the higher will be the level of employment, for a given real wage and level of investment.

Consider an upward shift in investment (the dotted line \( C+I' \)). Employment and output both increase, by a multiple of the increase in investment. Profits increase by the exact amount of the additional investment. This follows tautologically from the definitions \( Y=W+P \), \( E=C+I \); the equilibrium condition \( Y=E \); and the assumption \( W=C \). But there is more than mere tautology behind this. "Workers spend what they get; capitalists get what they spend." The point of this aphorism is that the direction of causality is different in the wage-consumption and profit-investment relationships. Wages are the source of consumption spending; workers have no income-yielding assets, so cannot sustain a spending stream unless they receive wage income. By contrast, capitalists, owning the means of production, are in a position to initiate spending, which then generates the profits to underwrite it. The multiplier depends exclusively on the slope of the real wage line; it is the reciprocal of the complement of the wage rate, \( \frac{1}{1-w} \). The aggregate "propensity to consume," based on psychology, has nothing to do with it. Profits are retained in the short run, and used for finance. Wages on the whole, are spent. The multiplier reflects the real wage and the characteristics of the property system in this simplified model.

Suppose, for some reason, business offered a level of employment different from the equilibrium. Can we say, on the basis of this analysis, what would happen? In orthodox theory, dynamics is notoriously tricky, requiring special assumptions, which often prove difficult to justify, about the pattern of market adjustment to disequilibrium. But here no special assumptions are necessary. All that is necessary for a rudimentary dynamic analysis has already been incorporated into the concept of the utilization function. Employers adjust their current production to their current sales, keeping inventory stocks constant.\(^7\) If the employment

\[ Y = \frac{a}{1 + \frac{c}{e}} \]

\[ P = Y - W \]

\[ \frac{1}{1-w} \]

\[ \text{Multiplier} = \frac{1}{1-w} \]

\[ 1-w \]

\[ \text{Wages on the whole, are spent. The multiplier reflects the real wage and the characteristics of the property system in this simplified model.} \]

\[ \text{Suppose, for some reason, business offered a level of employment different from the equilibrium. Can we say, on the basis of this analysis, what would happen? In orthodox theory, dynamics is notoriously tricky, requiring special assumptions, which often prove difficult to justify, about the pattern of market adjustment to disequilibrium. But here no special assumptions are necessary. All that is necessary for a rudimentary dynamic analysis has already been incorporated into the concept of the utilization function. Employers adjust their current production to their current sales, keeping inventory stocks constant.} \]

\[ \text{\(^7\) More complicated dynamics are easily developed following well-known models. Time lags could be introduced, movements could be assumed to generate expectations of further movement in the same direction, and so on, producing familiar patterns of inventory fluctuation. See R. G. D. Allen [1]. Kalecki [4].} \]
offered is below equilibrium, say at $N_1$, then the aggregate demand $C_1 + I$, will be greater than $Y_1$, the output corresponding to $N_1$. Inventory will be run down, and additional plant and equipment, idle initially, will be brought into operation, leading to the employment of more workers. Employment will rise to $N_2$, but at this level of employment the wage bill, and so consumption spending, is now greater. Hence, aggregate demand now equals $C_2 + I > Y_2$. Again, inventory will be run down and more idle plant and equipment will be brought into operation, with employment rising to $N_3$ and aggregate demand to $C_3 + I > Y_3$, and so on until the equilibrium point is reached.

Starting from an increment of investment, the resulting change in output is given by summing the series of the rounds of responding. The first term is given by the wage times the amount of additional employment, caused by the increase in investment: the second term by the wage times the first term, and so on. Hence, the additional consumption each time is the fraction of investment spending that goes to wages, which will be the wage rate times the amount of employment generated by the investment spending. Employment per unit output, $n = \frac{\Delta N}{\Delta Y}$, is the reciprocal of the productivity of labor, the slope of $Y = f(N)$. So, the multiplier series is,

\[
\frac{\Delta Y}{\Delta I} = 1 + wn + (wn)^2 + \ldots = \frac{1}{1 - wn}
\]

\[
\frac{\Delta N}{\Delta I} = n\left(1 + wn + (wn)^2 + \ldots\right) = \frac{n}{1 - wn}
\]

In the case where $n=1$, because employment and output are both measured as percentages of capacity levels, this becomes

\[
\frac{\Delta Y}{\Delta I} = 1 + wn + (wn)^2 + \ldots = \frac{1}{1 - wn}
\]

\[
\frac{\Delta N}{\Delta I} = n\left(1 + wn + (wn)^2 + \ldots\right) = \frac{n}{1 - wn}
\]

---

8 These two series are derived from the following sequences:

\[
\Delta Y_1 = \Delta I
\]

\[
\Delta Y_2 = \Delta C_1 = wn\Delta I
\]

\[
\Delta Y_3 = \Delta C_2 = wn\Delta C_1 = (wn)^2 \Delta I
\]

\[
\Delta Y_4 = \Delta C_3 = wn\Delta C_2 = (wn)^3 \Delta I
\]

\[
\Delta N_1 = n\Delta I
\]

\[
\Delta N_2 = n\Delta C_1 = n(wn\Delta I)
\]

\[
\Delta N_3 = n\Delta C_2 = n(wn)^2 \Delta I
\]

\[
\Delta N_4 = n\Delta C_3 = n(wn)^3 \Delta I
\]

\[
\text{etc.}
\]

\[
\text{etc.}
\]
Notice that the effects of the multiplier are concentrated in the consumer goods sector after the initial round of spending. It is there that employment and output increase.

At this point it is evident that it will not do to treat the economy as if it consisted of “one sector” only. Investment demand and consumption demand are directed to different parts of the economy and need to be treated separately. Accordingly, the preceding account of the multiplier needs to be restated. Define:

\[ n_i \text{ as the labor coefficient in investment goods} \]
\[ n_c \text{ as the labor coefficient in consumption goods} \]
\[ w \text{ as the real wage, assumed to be wholly spent} \]
\[ u \text{ as the ratio } \frac{w - z}{w} > 0, \text{ where } z \text{ is the level of spending which can} \]
\[ \text{be sustained by an unemployed worker, financed by borrowing and} \]
\[ \text{the Welfare State} \]
\[ a \text{ as the coefficient of replacement demand by the consumer goods} \]
\[ \text{sector for investment goods.} \]

Let \[ x = uw n_c \]
\[ y = uw n_i \]

Then when effective demand varies due to a change in investment, \( \Delta I \):
\[ \Delta C_1 = y \Delta I \]
\[ \Delta C_2 = xy \Delta I + ay^2 \Delta I = (x + ay)y \Delta I \]
\[ \vdots \]
\[ \Delta C_n = (x + ay)^{n-1} y \Delta I. \]

So the series will be:
\[ y \Delta I[1 + (x + ay) + (x + ay)^2 + \cdots] \]

which gives,
\[ \text{multiplier } = \frac{\Delta C}{\Delta I} = \frac{y}{1 - (x + ay)} = \frac{uw n_i}{1 - uw(n_c + an_i)} \]

This formula shows sectoral coefficients explicitly together with the intersectoral relationship. The principle impact will take place in the consumer goods sector; the effect on employment in investment goods will be secondary. It also shows the effect of welfare and unemployment compensation. The higher \( z \) is, the lower will be \( u \) and so the lower the multiplier. This is intuitively obvious; the change in spending as a result of moving from employed to unemployed, or vice versa is not so great.²

²Let \( m \) be the multiplier, and let \( A \) stand for \( wn_i \) and \( B \) stand for \( w(n_c + an_i) \). Then,

\[ \text{Change in spending as a result of movement from employed to unemployed is explainable in this model.} \]
Now suppose \( z = 0 \). Rewrite the formula and cross-multiply. We have:

\[
wn \cdot l = [(1 - w(n, \ a(n, \ l))] \Delta C.
\]

The left hand side is the additional consumption goods demand resulting from additional employment in the investment goods sector. The right hand side is the additional sales of consumer goods, minus the additional sales of consumer goods to workers in the consumer goods sector itself. minus the additional sales of consumer goods to workers newly employed in investment goods to produce additional replacements for consumer goods. In short, the equation reads:

the extra demand for consumer goods by investment goods workers = the extra supply of consumer goods (over and above the costs in consumer goods of producing this supply).

This is, therefore, a dynamic and more complex form of the famous Marxian "balancing condition" for expanded reproduction. [11]

In this formulation (as in the simpler version above) the multiplier no longer depends on anyone's "psychological propensity" to consume. The multiplier here is based upon structural and institutional features of capitalism. It depends first and foremost on the level of real wages, which reflects the division of income between pay and profits, or, more generally, between variable costs and the mark up. It further depends on the labor coefficients in the sectors and on the degree of technical dependence between them, and finally, upon the welfare policy of the State. Of course, if workers save a portion of their income this can easily be taken into account: workers will simply spend a fraction of their wage, rather than the whole of it. A more serious modification may be required to take into account the fact that it takes households time to adjust their spending habits when real wages change. But this would take the argument out of the realm of simple models.

It has already been argued that there are good reasons for supposing both money wages and prices to be unresponsive in many circumstances.

\[
\frac{dm}{du} = \frac{\Delta N}{(1 - u)} > 0,
\]

Keynes, in a letter to Beveridge, 28 July 1936, presents the multiplier in this light, rather than, as in the General Theory, depending on a psychological propensity. "Take the case of an increase in investment, say, the building of additional houses. The men who are directly employed in building the houses will have a higher income than before. They will spend this income on consumption... except when there is full employment, there is an elasticity of supply in the consumption-goods industries, and... if more men are employed in building houses, more men will also be employed in making things for the house-builders to consume... The additional men employed in the consumption industries will themselves consume more, so that we have a whole series of repercussions." [8: pp. 57-8]. Joan Robinson's account of the multiplier runs along these lines [13: pp. 15-22]. R. F. Kahn in his famous original article takes account of both inputs and the "stale" (welfare support of unemployed), but mistakenly believes, first, that an increase in employment will yield an increase in realized profits in the short run, and secondly, that there will be current spending out of these profits. See Kahn [5: pp. 11-12]. This erroneous perception of the role of profits in the short run pervades the entire Keynesian literature. By contrast, Kalecki was always clear that investment expenditure (and capitalist consumption) determined realized profits, which, therefore, exerted no influence on current spending.
to short-run changes in the degree of utilization. However, there are certainly plausible conditions in which changes in utilization would be accompanied by either price or money wage changes or both. Such changes can easily be incorporated in the analysis. Consider a decline in investment spending, causing unemployment. Suppose that producers have a tacit agreement not to "spoil the market," so that prices fall very little, but the labor market is highly competitive and the division of labor has proceeded far enough that workers are easily interchangeable appendages to the assembly line. So wages will fall while prices stay steady, or decline more slowly. To keep the argument simple, look at the 45 degree diagram in Figure 4. Initially investment was sufficient at \( I_1 \) to provide full capacity employment. It then shifts down to \( I_2 \) generating unemployment \( \Delta N_1 \). This leads to a fall in the wage rate, a change in the slope of the wage line, from \( w_1 \) to \( w_2 \), lowering the wage bill, and creating additional unemployment \( \Delta N_2 \), entirely concentrated in the consumer goods sector.

![Figure 4](image)

These two processes will take place together. The decline in investment spending will create initial unemployment, which will then lower the money wage (and so the real wage) a certain amount; these two effects will together lower spending on consumer goods, creating more unemployment which will then further lower the money wage; the two effects then causing a further reduction in consumer spending, and so on.

Formally (using the simpler version of the multiplier),

\[
\Delta Y_1 = w_1 \Delta N + N_1 \Delta w + \Delta w \Delta N \text{(ignoring the last term)} = w_1 n \Delta I + N_1 x n \Delta I = (w_1 + xN_1) n \Delta I, \text{ where } w_1, N_1 \text{ is the initial wage--employment position, } n \text{ is the (aggregate "one-sector") labor coefficient, and } x \text{ is the fraction by which the wage changes when employment changes, assumed to be constant through the process and to be the same for increases and decreases.}
\]

\[
\Delta w = x \Delta N = x n \Delta I
\]
\[ \Delta Y_2 = [(w_1 - \Delta w) + x(N_1 - \Delta N)](w_1 + xN_1) n^2 \Delta I \]
\[ \Delta Y_3 = [(w_1 - 2\Delta w) + x(N_1 - 2\Delta N)][(w_1 - \Delta w) + x(N_1 - \Delta N)](w_1 + xN_1) n^2 \Delta I. \]

and

\[ \Delta Y_k = [(w_1 - (k - 1)\Delta w) + x(N_1 - (k - 1)\Delta N)]n\Delta Y_k - 1 \]
\[ = [(w_1 - xN_1) - (k - 1)(1 + x)\Delta I]n\Delta Y_k - 1 \]

By taking \( k \) sufficiently large, the last term in brackets can be made zero; hence \( \Delta Y_k = 0 \). The series is therefore a finite sum of \( (k - 1) \) positive terms.

If cut-throat competition caused prices to fall, while money wages remained sticky in the short run, then the real wage would rise when investment demand fell. Such a rise would dampen, or, in the extreme case, reverse the effects of the decline in investment spending on employment and output.

And this brings out a very important proposition; a rise or fall in the real wage brings about a corresponding rise or fall in the level of employment in the consumer goods sector, together with a secondary effect (through the coefficient, \( a \)) on employment in investment goods. In the theory of effective demand, employment and the real wage vary together.

This means that existing unemployment (at least in consumer goods) can always be eliminated by a large enough increase in the real wage. Suppose present investment is too low to generate full capacity employment, yielding instead employment, \( N_0 \). (Figure 5) Draw a line, \( IF \), from the investment intercept on the \( Y \) axis to the full employment point \( F \). Then the real wage required for full capacity employment is the wage line parallel to the dotted line \( IF \).

---

**Figure 5**

---

11 The contrast is worth spelling out: in marginal productivity theory a lower real wage implies more employment — employers are stimulated by cheap labor. In the theory of effective demand a rise in the real wage implies a rise in consumption sales — employers are stimulated by inventories running down.
Another way to put the same point: the level of investment necessary to produce full capacity employment is a unique and inverse function of the real wage. Given a factory system with a fixed capacity, the higher the real wage, the lower the level of investment and government spending needed for full capacity performance.

A qualification is necessary, however. The increased employment $N_F - N_0$, will chiefly take place in the consumer goods sector, with a secondary effect in investment goods due to increased demand for replacements of worn out consumer goods equipment. This secondary effect may not be very large, so the higher real wage can assure full capacity employment only in the consumer goods sector.

Notice, however, that the profits of capitalists are the same absolute amount at both levels of employment, $N_0$ and $N_F$. Since the capital stock is given, whether used or not, both yield the same average rate on invested capital. Profit is determined by investment spending (plus exports, government deficit spending, and capitalist consumption, all neglected here) not by the level of the real wage. In the short term, changes in the real wage have no effect at all on realized profits.\(^{(12)}\)

The argument can be developed a step further by dropping the convention of measuring output and employment as percentages of capacity. Measure them instead in money value and man-hours. Then for given levels of investment and real wages, the lower the productivity of labor, the higher will be the level of employment. On the diagram, (Figure 6), the

![Diagram](image)

**Figure 6**

\(^{(12)}\) Might not a rise in the real wage induce a decline in investment spending, and in this way bring about a decline in profits? This is certainly possible, but the relationships involved go beyond the simple model of this paper. Investment spending, taken as exogenous here, depends on prior investment decisions, which the firm takes in pursuit of growth. The firm will be eyeing potential markets, new products, new technical processes, and balancing this against further expansion in its existing markets. Investment decisions are decisions to build a certain plant, embodying a certain technical process, to produce a certain product, designed (literally) to sell in a certain market. The level of real wages expected both during construction and later in operation will obviously be relevant to such decisions, as costs; and the general level of real wages will be relevant in terms of the expected size of the market. Once investment decisions are taken, there is the further question of how rapidly to implement them. Investment spending can be speeded up or slowed down, depending on the phase of the trade cycle, the availability and cost of finance and the political and social climate. Here again, the level of real wages will be important, but the effects will likely be different in the different sectors.
dotted line shows a lower level of productivity, with the corresponding high level of employment, *ceteris paribus*. Employers will not employ more workers than they need to produce the output they can sell. So the interests of workers and employers are clearly opposed, since employers will not wish to pay high wages on long run and on competitive grounds. (What is true for the aggregate need not hold for the individual taken separately.) The interests of the workers and employers are clearly opposed in the short run.

An even stronger proposition can be demonstrated with this same diagram, one which is relevant to current policy proposals. Assume that investment demand is given exogenously. Then, if wages and productivity are increased by the same percentage, the equilibrium level of output will be higher and that of employment lower. This can easily be seen on the diagrams in Figures 7 and 8. The angle through which the productivity line rotates, $YOY'$, equals that through which the wage rotates, and that, in turn, determines the swing of the effective demand line, $EIE'$. Increased productivity reduces equilibrium employment from $N_0$ to $N_1$. The increase in the real wage raises it to $N'$. The proposition will be true if $N_0 - N_1 > N' - N_1$. Let the angle be $\alpha$: a line segment, $0Y'$, cutting it from below will always have a projection in the horizontal axis, less than that of a line segment, IE cutting it from above. The increase in productivity will *always* lay off more workers than the spending of the higher wages will re-employ. Think of it this way: each worker produces more than he consumes, $Y > w$: hence, $xY > xw$, where $x$ is the percentage increase in productivity and the real wage. So $xY > xwN$: the absolute increase in output is greater than the absolute increase in consumption. Since $E' = C + xc + I < Y + xy$, employment must fall. The idea that raising wages *parti passu* with productivity is somehow "neutral" or "fair" will not stand inspection.

![Figure 7](image1.png)  
![Figure 8](image2.png)
(It is important to remember two caveats: first, that the propositions demonstrated here, with the exception of those concerning the multiplier, are all comparative static. Secondly, I am neglecting the important relationships between finance and investment spending, and between aggregate spending and finance. So, investment spending is taken as exogenous; and when the text speaks, e.g. of "the increase in productivity" or of a "rise in real wages," this should be understood not as an actual change in historical time, but as a comparison between two equilibrium positions, in one of which productivity or wages is higher, with everything else the same.)

**Marginal Productivity**

In short, the **theory of effective demand** is a theory of employment in the short run. But **neo-Classical theory** is committed to the theory of marginal productivity, according to which an increase in employment can come about only if the real wage declines. And Keynes explicitly adopted this position in *The General Theory*:

... in general, an increase in employment can only occur to the accompaniment of a decline in the rate of real wages. Thus I am not disputing the vital fact which the classical economists have (rightly) asserted as indefeasible. In a given state of organisation equipment and technique, the real wage earned by a unit of labor has a unique (inverse) correlation with the volume of employment. Thus, if employment increases, then, in the short period, the reward per unit of labour in terms of wage-goods must, in general, decline and profits increase. This is simply the converse of the familiar proposition that industry is normally working subject to decreasing returns in the short period during which equipment etc. is assumed to be constant... [7: p. 17]

Keynes could hardly have been clearer. He accepts the whole of marginal productivity theory — short period decreasing returns, the real wage equal to the marginal product, and any increase in employment implies a decline in the real wage. However, he does say that, "equipment etc. is assumed to be constant." But if factories and their equipment are given, then we are considering the question of utilization, and there is no reason to suppose that returns diminish. So long as labor's marginal product is greater than the real wage, it will pay business to employ labor in proportion to the demand for goods, up to capacity. The question then arises, are there no forces acting to equate the real wage and the marginal product?

In the true neo-Classic case, the short period is defined by the amount of capital, which remains constant. Every point on the production function is efficient; labor and capital are combined in the technically optimal manner. Hence each point represents a complete adaptation of the capital equipment; each point has its own unique history. In this case the neo-Classic supply and demand curves for labor can be used to define "full employment," and "involuntary unemployment." will then exist
only when labor is "off" its supply curve. But such a conception is wholly
unsuitable for analyzing short-run changes in employment. Moreover this
is precisely the conception of the production function which the capital
telegraph controversy has shown to be in general untenable.\textsuperscript{13} Short-run
changes in employment must be analyzed in terms of changes in capacity
utilization.

First there is the question of the shape of the utilization function. If
returns are constant or increasing then it is impossible for the real wage to
equal the marginal product of labor. Employers will hire labor so long as
the real wage is less than or equal to the marginal product, but if addi-
tional employment increases the marginal product or leaves it unaffected,
there will be no tendency for the two to be equated. Diminishing returns
to factory utilization is therefore the only case consistent with orthodoxy;
though perhaps the least plausible empirically.

Even this implausible case gives unacceptable results. Suppose the real
wage is given, as above, and that a certain amount of investment spending
is scheduled. Suppose that this gives an effective demand that cuts the
employment function in two places as drawn. (Figure 9). At the lower
point, \(N_1\), the real wage is less than the marginal product, at \(N_2\) they
are equal, and at \(N_3\) the real wage is greater than the marginal product. \(N_3\)
should therefore be the profit maximizing point, and it is indeed the point
at which potential profits, output minus wage costs, are the greatest. But
realized profit is equal to investment, and so will be the same at all these
levels of employment. At \(N_1\), however, or anywhere between \(N_1\) and \(N_3\),
there will be unsold output; only at \(N_1\) and \(N_3\), are output and
expenditure just balanced. But there is a significant difference between
the two points: just below \(N_1\) demand is greater than output, while just
above it output exceeds demand. So if employment is near \(N_1\), business
will be motivated to raise or lower it in the direction of \(N_1\). By contrast, if
employment is just below \(N_3\) output is greater than demand, leading

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure9.png}
\caption{Figure 9}
\end{figure}

\textsuperscript{13} For a recent account of the controversy, See Laibman and Nell [9]. Also for a full history,
Harcourt [2].
business to lay workers off; just above \( N_1 \), demand exceeds output, leading to further expansion, and a still wider discrepancy. In short, \( N_1 \) is a stable equilibrium point, and \( N_2 \) is unstable.

This is the true - and only - significance of the marginal productivity doctrine in this case. At a level of employment where the real wage is less than the marginal product, the corresponding aggregate effective demand function will intersect the output curve in a stable point, when the real wage is greater than the marginal product, the point of intersection will be unstable. This follows very simply from the fact that as long as investment spending is autonomous, the slope of the effective demand function is given by the real wage. Hence when the real wage is less than the marginal product, the demand function cuts the output curve from above - and the point is stable.

In its microeconomic form, marginal productivity is a theory of employment: employers face a given wage and hire workers until their marginal product declines to the level of the wage. But here, given the level of investment, and given that employment will be adjusted until output equals demand, for the theorem to hold the real wage would have to adjust to the profit maximizing marginal product. When there are diminishing returns to utilization, given any level of investment there is always a real wage such that the effective demand line will just be tangent to the output curve. (For each level of investment, e.g., \( I_1 \), measured on the vertical axis, Figure 10, one and only one line can be drawn tangent to the curve, given the latter's convexity and continuity. The required real wage, \( w_1 \), will then be the ray from the origin parallel to the tangent line.) Given any real wage, e.g., \( w_2 \), there is always a level of investment, \( I_2 \), such that the effective demand line will be just tangent to the output curve. But there do not seem to be any economically plausible pressures that would tend reliably to move either the real wage or investment to the level corresponding to the tangency position.\(^{14}\) If the wage were less than the marginal product it would have to rise, even though unemployment existed. Alternatively, investment would have to rise, even though there existed unused capacity. Nevertheless, suppose that investment spending were increased just enough so that the effective demand line became tangent to the output curve at \( N_2 \), on Figure 9. The real wage would then exactly equal the short-run marginal (utilization) product. Now consider some simple dynamics. At levels of employment just below \( N_2 \), demand

\[\text{\(^{14}\)What if prices were allowed to change? If excess capacity implies lower prices, unemployment should imply lower money wages. If the effects are the same, the real wage remains unchanged. But why should one effect systematically outweigh the other? There seems to be no clear answer. If prices were more flexible than money wages, then excess capacity would lower the real wage and so capacity utilization; if money wages were more flexible, the real wage would rise, reducing excess capacity. Both effects were discussed earlier along with the multiplier, but neither tend to move the real wage towards the marginal product. To keep things simple, however, it has seemed best to assume that neither prices nor money wages change (or are different) when demand varies at levels below full capacity.} \]
> output, so employment will move towards \( N_2 \); but at levels of employment just above \( N_2 \), demand will also be greater than output, leading inventories to run down, inducing business to hire more to meet the demand. The fact that output in the aggregate does not keep pace does not reduce profits, which remain equal to investment. Inventories, however, are run down increasingly, until output hits capacity. At that point, output can no longer be increased further, and demand (plus the demand by business to restore inventory) pushes up prices. This reduces the real wage, until the effective demand line has swung down until it just meets the output curve at full capacity.

![Graph](Figure 10)

At this point, the factory system is fully utilized and no more labor can be hired; there are no places on the assembly line. So, given investment spending, demand cannot increase further. But if for any reason employment should temporarily fall below the capacity level, output will then be greater than aggregate demand, leading to overstocking and inducing business to lay workers off. The consequent reduction in worker spending will reduce demand more than the layoffs will reduce output, and the recessionary gap will first grow, and then diminish as employment is progressively reduced, finally reaching stable equilibrium at the comparatively low level \( N_0 \), below the level at which the wage equals the short-run marginal product. (Figure 11)

![Graph](Figure 11)
The theory of effective demand, then, is a theory of employment of labor in the short period, where labor is employed to operate equipment of fixed capacity, and to produce an output adjusted (with allowance of maintaining inventory) to the current level of demand.

But although this is a theory of employment and of aggregate output, in Keynes' sense, far from being opposed to the theory of distribution (as Keynes thought it had to be) it is at the same time a theory of distribution. Investment (and related kinds of spending) determines profits, while wages determine consumption. At capacity, excess effective demand will cause prices to rise relative to money wage rates until profits generate the appropriate savings to offset investment. It is a short-run, and oversimplified theory, but it establishes the connections between effective demand and the distribution of realized income at a very general level. It allows us to see the error in the orthodox theory quite plainly: orthodoxy assumes that employment will settle at the point where the difference between output and wage cost is at a maximum, regardless of whether the resulting product can be sold. The assumption is plain: whatever is produced will be sold because supply creates its own sufficient demand. By contrast the theory of effective demand determines employment at the level where output will be exactly matched by demand.

In this light the argument seems simple and obvious, yet it runs counter to almost all the conventional wisdom of both monetarists and Keynesians. The theory of effective demand really supports neither side of the current dispute between the two branches of orthodoxy although it is unquestionably closer to the Keynesians. But, it does show quite clearly why neither can present an adequate account of their relationships between inflation and unemployment: both sides start their analysis by assuming marginal productivity theory.

**Wealth Effects**

The orthodox reply will be that everything presented so far depends on "sticky" money wages and prices, i.e., on market imperfections. With fully flexible money wages and prices, the emergence of a shortage of effective demand would lead to an equiproportional fall in both, resulting in a higher real value of cash balances. This improvement in household and business wealth positions would entail higher spending, overcoming the demand shortfall. Strictly speaking, the argument is a comparative static one; it states that there always exists a low enough level of prices and money wages to raise the value of cash balances enough to provide a full employment level of spending. (It is generally agreed that even if money wages and prices were flexible downward, no dynamic argument could be made, since declining wages and prices are likely to set up adverse expectations.)

Consider an economy with given plant and equipment, financed by
some given pattern of borrowing. (Orthodox theorists often assume for simplicity that all capital is raised by selling bonds.) Given expected money outlays and revenues and the requirements of debt servicing, businesses and households will hold certain nominal cash balances. Now suppose that the level of consumption and investment is insufficient to employ the economy's full capacity at the initial level of money wages and prices. Suppose money wages and prices were lower, but that everything else — the real wage, plant and equipment; debt structure, and nominal cash holdings — remained the same. Could we expect a higher level of spending?

(It cannot be argued that because the nominal value of plant and equipment is lower, the nominal amount of debt will also be lower: because, then, by the same token, nominal cash balances would also be less. Holding cash is an alternative to paying off debt; if the one is fully adjusted, the other must be also.)

The orthodox answer is that the higher real value of cash balances, making households and businesses wealthier, will cause the psychological propensities to consume and invest to shift upwards. The increased wealth does not yield an increased flow of purchasing power; it merely stimulates economic agents to spend more out of their existing and unchanged flows of income.

Two problems spring to mind at once. Since the present model assumes that all wage income is spent and that only wage earners consume, there is no room for the "wealth effect" in consumption. Room could be made by allowing worker saving or introducing a salaried and parsimonious middle class, but this just pushes the problem back a step. What happens when the "wealth effect" has done its work, stimulating consumption until saving disappears, and still \( C + I < \) full capacity output?

The second obvious point concerns installment buying and resale value. Plant and equipment, houses, automobiles, and many consumer durables are commonly financed by vendor's mortgages. An important component in the decision to buy is the expected resale or second-hand value, at the end of a certain period, with which the tail end of the mortgage will be retired. Lower money prices will affect this calculation, and would require a curtailment of current spending.

But the chief defect in the doctrine of the "wealth effect" is that it ignores the increase in the real burden of debt. When money wages and prices both decline, so does the absolute size of gross profits. This can create problems in debt servicing, and lead to insolvency; if it goes far enough, it will reduce gross profit below the level of interest plus principal. Instead of deflation leading to expansion, deflation leads to bankruptcy, which, in practice, of course, it always has. (If cash balances are run down to meet the demands of debt servicing, bankruptcy may be avoided — this period! — but all possibility of expansionary stimulus is gone.)
So the orthodox position has to be that the burden of indebtedness can be reduced by rolling over the original bonds at a new, lower rate of interest. The rate of interest will be lower because the transactions demand for money is lower, owing to the lower level of wages and prices. In rolling over the debt, the volume of new bonds demanded will be exactly the same as the old bonds being retired, hence the operation, skillfully managed, finances itself and should have no effect on interest rates.

The trouble is obvious as soon as one looks at the deal with the eyes of a banker. A firm has machinery and plant initially worth $x and mortgaged for that amount. Prices have fallen, however, and the new value is $x/y; where $y > 1$. It wishes to pay this mortgage off now, by issuing bonds or borrowing $x$ at a lower rate. But the security it has to offer for this new loan is now worth only $x/y$. Why should a bank or any sensible financier lend against such inadequate security? The only possible justification for lending against inadequate security is that the rate of interest offered is high enough to warrant the risk. But the orthodox position requires that the new rate of interest be lower, even though the risk is manifestly greater!

Flexible money wages and prices provide no escape from the conclusions of the theory of effective demand.

Sectors

Enough has been said already to show the importance of separating the economy into sectors, and it is time to develop this explicitly. First we will exhibit a position of long-run equilibrium, in which the wage rate and the rate of profit are equal in both sectors at full capacity. Employment in each sector, $N_1$ and $N_C$ (Figure 12) is measured in manpower along the horizontal axis from the origin to the right; capital invested for each sector, $K_1$ and $K_C$ is measured in money value from the origin to the left along the same axis. The vertical axis reading upward in each case measures output in money value, $Y_1$ and $Y_C$, and reading down from the origin it measures capital again.

![Figure 12](image_url)
The sectors are drawn with consumer goods assumed labor intensive and capital goods capital intensive. The slopes of the lines \( K_1 Y_1 \) and \( K_C Y_C \) give the respective output-capital ratios, those of the lines \( K_1 N_1 \) and \( K_C N_C \) the capital-labor ratios, and those of \( O_1 F_1 \) and \( O_C F_C \) the labor productivities. Angles \( W_1 O_1 N_1 \) and \( W_1 O_C N_C \) represent the wage rates, equal in both sectors, and angles \( P_1 K_1 O_1 \) and \( P_C K_C O_C \) the profit rates, also equal. Lines \( F_1 P_1 \) and \( W_1 O_1 \) are parallel, as are \( F_C P_C \) and \( W_C O_C \). Hence, \( O_1 P_1 \) and \( O_C P_C \) are the respective amounts of profit in the sectors. If all, and only, this profit is invested then the two sectors will grow uniformly.

The balance between the sectors is shown by line \( W_1 P_C \), which expresses the fact that profit in the consumer goods sector equals the wage bill in investment goods.

To see the connection of this with long-run pricing, consider what happens if the real wage were to settle at a level equal to half the value shown. The new potential rates of profit at full capacity operation in both sectors are shown by the slopes of the dotted lines \( K_1 P_1 \) and \( K_C P_C \). Clearly the fall in the real wage raises the potential rate of profit in the labor-intensive sector above that in the capital-intensive one. What is required, as is well known from Ricardo, Marx, and Sraffa [14] is a higher relative price of capital to consumer goods, lowering the potential rate of profit in consumer goods and raising it in capital goods. But if we look at this in terms of monetary magnitudes in the short run, it is not at all clear how this might come about. If the investment goods sector operates at full capacity, its wage bill will now be \( W_1 \); hence, profits in consumer goods will be \( O_1 A \), and the line \( AB \) will give the total demand for consumer goods. This intersects the output-employment line \( O_C F_C \) at a level of employment, \( N_C < N_C \). There will, therefore, be unemployment and excess capacity in consumer goods and the realized rate of profit will be shown by line \( K_C A \), less in slope than \( K_C P_C \), and far less than \( K_1 P_1 \). The price of consumer goods and/or money wage rates may fall, liquid capital may shift, investment spending may fall off, and investment plans may alter. These questions are beyond the scope of this paper. The point is that with given plant and equipment and differing sectorial capital-labor ratios, a change in the real wage both renders the potential profit rates at full capacity unequal and also undoes the short-run sectoral balance. The restoration of full equilibrium requires that both conditions be corrected.

Here our concern is with the short run. Putting aside the problem of unequal rates of profit, consider the relations between the sectors in terms of effective demand, using the core of the same diagram. (Figure 13)

The discussion which follows concerns comparisons of equilibrium positions before and after changes in the value of some variable or variables, in spite of the fact that the language may appear to refer to movements in variables. To study movements — dynamics — we would have to...
take into account the effects of movement upon expectations, etc., and derive the explicit time paths of the variables, an enterprise beyond the scope of this paper.

Initially gross investment demand is at capacity and with the given real wage, employment in consumer goods is \( N_C \) and output \( Y_C \), both also at capacity. Now let investment demand shift down to \( Y'_1 \), so that equilibrium employment falls from \( N_i \) to \( N_i' \). This reduces the wage bill in investment goods from \( W_i \) to \( W'_i \), and consequently lowers profit, output, and employment in consumer goods. But the decline in employment in investment goods depends only on the labor coefficient there, whereas the decline in employment in consumer goods depends both on the productivity of labor in consumer goods and on the real wage. The higher the productivity of labor, given the real wage, the smaller will be the decline in consumer goods employment for a given fall in investment demand. Given the productivity of labor, the higher the real wage, the larger the decline in employment. If investment goods is capital intensive and has a therefore, the higher productivity of labor, then, especially in high-wage economies, fluctuations in investment demand may cause greater changes in employment in consumer goods than in the investment sector itself.

Now consider a change in the real wage, for a given level of investment spending, using the same diagrammatic apparatus. (Figure 14) We have already seen this above, but it is worth a closer look.

Gross investment demand stays steady at \( Y_C \). At the initial level of the real wage both sectors are fully employed. (This is for simplicity only; there could be unemployment in both, which might cause money wages to fall further than prices, bringing about the lower real wage.) Now consider
a lower real wage. In investment goods, profits are now $P_i > P_L$, while
employment and output remain the same. In consumer goods, however,
equilibrium profits decline from $P_L$ to $P_L'$, output falls to $Y_L$, and unem-
ployment $N_L - N_L'$ emerges. But the total profit remains the same, for
the increase in investment goods profit, $W_L - W_L'$, exactly equals the
decline in consumer goods profit, $P_L - P_L'$. Just the reverse would happen
with a rise in the real wage. Thus in investment goods, profit varies
inversely with the real wage, while in consumer goods, it varies directly —
a proposition which should lead the capitalists of the two sectors to take
different approaches to economic policy.

The decline in employment and output in consumer goods is brought
about by two causes, represented respectively by the changes in the
intercept and the slope of the consumer goods total demand line, AB. The
new intercept shows the decline in the demand for consumer goods by
capital goods workers due to their lower wages; the new slope shows the
lower wages (and no consumption) of workers in the consumer goods
industry itself. It is clear that high real wages benefit all parties in the
consumer goods sector in the short run.

Now let us turn back to the question of productivity increases. Let
investment spending be given. Suppose first, implausibly, that produc-
tivity rises at the same rate in both sectors, and that the real wage is raised
by the same percentage, a supposedly desirable and non-inflationary
increase. In the investment goods sector, the rise in productivity will
reduce employment, leaving output unchanged, but the rise in the wage
will exactly offset the reduction in employment, leaving the wage bill
unchanged. Profits are therefore also unchanged. Since the investment
sector’s wage bill is unchanged, profits in the consumer goods sector will
be the same. But output in consumer goods will be up and employment
down, as can be seen by inspecting the diagram, which is the same as that
considered for the economy as a whole. (Figure 15)

![Figure 15](image)

Again assume that investment spending is fixed. Suppose now that
productivity increases faster in investment goods than in consumer goods,
but that wages are raised by the average of the increases in the sectors,
weighted by their sizes. Profit will increase in the investment goods sector, because the decrease in employment will be proportionally greater than the rise in wages. Hence the wage bill will diminish. So, by the same token, profits in the consumer goods sector will be down. The total profit will be the same, of course. If productivity increased faster in consumer goods than in investment goods, profit would be squeezed in investment, and the increased wage bill there would raise profits in consumer goods.

Now consider what happens when, with capacity fixed and no technical progress, total demand for investment exceeds capacity. According to the "Cambridge" theory of distribution, when aggregate demand is greater than capacity, prices will be higher relative to money wages (excess demand for goods does not imply excess demand for labor, since the demand for labor is limited by the number of places on the assembly line), so profit will be larger, providing the saving to balance the investment, i.e., reducing consumption demand. The problem is that while the price of investment goods will be bid up, which will increase profits in that sector, additional profits do not mean additional capacity. The demand still cannot be met. So long as the money wage and the price of the consumer goods are unchanged, no real adjustment is possible. However, the higher price of investment goods raises the cost of materials and replacements in consumer goods. Hence, the price of consumer goods must go up. With given money wages, this means a decline in the real wage. Hence, output and employment decline in consumer goods. So the demand by the consumer goods sector for materials and replacements from investment goods will be lower, and it is this which provides the free capacity to meet the initial excess demand. The adjustment takes place by creating inflation in investment goods, which is passed on to consumer goods, where it is transmuted into recession, which in turn frees capacity in investment goods. A higher price of investment goods is sufficient to balance saving and investment in value terms; but higher profit by itself does not mean more serviceable capacity. Nor does a lower level of consumption, for that creates excess capacity in consumer goods, the wrong sector. It is the decline in the consumer goods sector's demand for the products of the investment goods sector which finally frees usable capacity.

The purpose of this paper has been to present the basic elements of the theory of effective demand in a manner making it possible to compare and contrast these ideas with those of orthodox theory. The detailed argument could be carried much further, but the basics should be clear by now. The post-Keynesian approach takes a realistic view of the economy.

---

15 Too much should not be claimed. This paper has abstracted from the determinants of investment. To complete the short-run analysis it will be necessary to provide a theory of finance which will explain the willingness and ability of business to carry through its spending. Both willingness and ability may well be affected, for example, by the shifts in profits just described. But short-run analysis cannot stand alone. It must fit into a framework which explains (as much as possible!) the long-run expansion and structural change of capitalism. Thus, the present argument must be complemented first by a theory of investment finance, and secondly by an account of investment decisions which fits into the theory of growth and development.
and society. It starts from the reality of social classes and industrial sectors. There are rigidities and they have important consequences. Property and property incomes are hedged about by rules which make the pattern of payment different from that of wage income. When these matters are taken into account, a theory of employment in the short run emerges which stands in flat contradiction to the neo-classical theory of marginal productivity. It also suggests promising new ways to account for "stagflation" and other contemporary problems. The next step is to look at the evidence.

References