The Economic Journal of Nepal,Vol.10,No.1,Issue 37,January-March 1987 (c) DOE-TU

General Economic Theory and the Emergence of Theories of Economic Development

KENNETH J. ARROW*

I am not, as you all know, a specialist in economic development, and doubtless most of you here have more knowledge of the problems of developing countries than I do. I intend here a modest aim. It is to review selectively some of the early formulations of the theory of economic development during the period 1945-1960 approximately and show how they related to developments in general economic theory at that and somewhat earlier periods. I am referring particularly to developments in neoclassical economics as it evolved under the pressure of the problems of developed countries. I also refer to that offspring of neoclassical theory which threatened to eat its father, Keynesian theory, also a response to the conditions of the advanced world. These had a strong and inevitable influence on thinking about economic development.

Let me briefly characterize neoclassical economics, as I understand it. It is rather a general viewpoint than a highly specific theory. It is constantly evolving. Neoclassical economics is an approach or mode of thinking which assumes optimizing behaviour by individuals together with an all over coherence of the economy in the form of balancing supply and demand. From the view point of application, what is relevant is a general tendency to balance rather than a moment-by-moment clearing of the market.

The charge is sometimes made that neoclassical economics is completely static and therefore unsuitable for analyzing change and growth. This charge cannot be maintained. Sir John Hicks, Gerard Debreu, and I, among many others, have shown how the future and even uncertainty can be brought into general equilibrium analysis, with suitable understanding and sophistication. Recent research has even extended the analysis to the role of information in the economic system, but this topic takes us beyond the period to be discussed.

Another recurrent theme is that neoclassical economics tends to the study of small-scale resource allocation problems rather than the grand themes which classical political economy is supposed to have displayed.

Nobel Laureate Dr. Arrow is currently Professor of Economics at Standford University. This article is based on his Presidential Address to the Eighth World Congress of the International Economic Association held in New Delhi in December 1986.

Personally, I believe that classical economics has the same fundamental world view point as neoclassical but is much less developed analytically. For example, technological unemployment in the long run is certainly a fallacy, but Ricardo's struggles with the problem in his chapter on machinery can only be described as confused. Consider also the meaning of gains from trade. Though Ricardo of course believed in and emphasized them, he had no way of explaining them, since in his system foreign trade represented an equal exchange of "values". He had to make some vague allusions to something different from value, which he called, "riches"; this is simply, "utility", in later language.

I want to relate some of the "grander" themes of development theory in its formative period to neoclassicial paradigms, including the failures of some paradigms to apply. I note that there are fewer and fewer of these grand theories and more and more highly specific applications. The neoclassical insight lends itself to this kind of analysis of specific situations, for example, the effects of a change in tariffs or agricultural subsidies. Indeed, the tendency to the specific is one of the advantages of the neoclassical approach. Grand visions do not lend themselves to useful policy. Following Ricardo or Marx leads one either to urge free markets everywhere or to say that there is no solution short of revolution.

It is useful to recall the state of economics and the pre-occupations of economists in the period before and after World War II and during the great era in which the developing countries achieved independence. It was at this time that the modern economics of development took its form. To over simplify, the main concerns of economists were (1) the ways in which the economy deviated from competitive behaviour and competitive efficiency, and (2) unemployment. Let me take these in reverse order.

Unemployment in the 1930's was too palpable and prolonged a phenomenon to be conveniently ignored by economists. In many ways, the presence of unemployment, of a failure to clear one of the most important of all markets, is a serious critique of neoclassical economics, with its emphasis on markets as a coordination of behaviour. Of course, Keynes's General Theory provided an intellectual representation which gave economists a tool for analysis.

From the view point of development economics, Keynesian theory gave encouragement to models in which labour played no constraining role. In particular, dynamic Keynesian models emerged, associated with the names of Harrod, Domar, and later, Joan Robinson. It was the unemployment problem of the developed countries that was under consideration, but these theories provided a major stimuls for development planning. They were convenient analytic tools, and there was a factual correspondence in that many of the developing countries exhibited large scale urban unemployment and atleast a perceived under-utilization of labour on the farms.

An intellectual problem remained: why is there unemployment in either developed or underdeveloped countries? The neoclassical (or, I

would maintain, even classical) argument that competition should drive wages down to eliminate unemployment are not on the face of it ridiculous. Policy-oriented economists and those whom they advised may at one time have regarded this kind of question as one for theorists to amuse themselves with. In the analysis of both developed and developing countries it was assumed that wages would remain constant as demand rose, atleast until a period of full employment (in developed countries) or labour scarcity (in developing countries) was achieved. The passage of time and better observation have shown the fallacy of this assumption if taken literally. Wages do rise even when the reservoir is not empty. A deeper understanding of the nature of unemployment still eludes us. Some version of the efficiency wage theory, with different interpretations for developing and developed countries, is the currently most coherent explanation of unemployment, but it is so far not very useful.

The presence of unemployment of course presents an opportunity for growth by providing a resource with zero shadow price, and indeed that was the assumption of even quite sophisticated planning models.

My second theme, you will recall, was that economic thought from about 1930 or so was much concerned with imperfections in competition and, related to that, the extent to which the actual market falls short of an optimum however defined. It has long been traditional in neoclassical economics to study market failures and defend state intervention as a way of overcoming them. The topic is not absent from Walras and is given considerable prominence by Marshall. It is however Marshall's student, A.C. Pigou, who presented the most complete catalogue of market failures and externalities, after his earlier efforts on faulty foundations were corrected by Allyn Young and Frank H. Knight.

Market failure and indeed any departures from Pareto optimality share with unemployment one important characteristic: there is the apparent possibility of a large gain in welfare, an impossibility if resources are already optimally allocated.

It is very frequently asserted that, in developing countries the market fails to achieve equality of wage rates in different sectors of the economy, the dual economy hypothesis. This doctrine seems to have been first expressed in a general form by Paul Rosentsein-Rodan, while discussing the economic development of southeastern Europe. When wage rates differ, marginal productivities differ, and there is room, perhaps considerable, for increases in welfare.

The simplest form of the idea is that farm wages or marginal productivities are lower than industrial or urban wage rates. Atleast three different causes are possible; all can be operating. (1) Since the members of a farm family share their receipts, the individual receives average productivity on the farm and only marginal productivity in the city. To induce mobility, urban wages would have to equal average farm productivity. Put another way, the ownership of capital on the farm is not transferred on a market but linked to family membership and participation in the farm work. (2) The transaction costs of migration and of information about the urban sector and the uncertainty of the move constitute a

second gap between rural and urban wages. (3) A third gap could be the political and collective economic power of the urban workers.

Instead of the gap between agriculture and industry, the dualism is sometimes found between traditional and modern sectors, in recognition of the importance and difficulties of the modernization of agriculture.

The explanatory power of this hypothesis is that it locates the cause of international differences in income levels in the failure of markets and limitations on mobility between them. The policy implications are ambiguous. The most direct implication would be to lower the institutional barriers and let the free play of the market improve resource allocation. But usually, the imperfection is taken as given, and policy has been oriented toward offsetting the imperfection with measures to subsidize and protect the industrial or, more generally, the modern sector.

The extreme version of the dual-economy hypothesis has a zero shadow wage for labour and thus is another form of the unemployment hypothesis. A constant real wage is needed in the urban sector. There is a striking and not accidental parallel to early Keynesian economics, where wagers would be constant until full employment is achieved, and, of course, the shadow wage is zero. In both spheres, those of developed and developing countries, more realistic analysis suggests that wages will rise as more workers are drawn into employment, in the Keynesian case, or into the modern sector, in the dual economy. The marginal productivity of farm labour is far from zero, as Theodore Schultz has shown.

No doubt many developing countries show dual economies. But the question remains, can a significant part of the international differences in income levels be attributed to dualism? Calculations of the production possibility curve for two-sector models, even with large distortions, suggest relatively small losses in industrial output for a given level of agricultural output. Further, the presence of large-scale urban unemployment or underemployment in a number of developing countries suggest that immobilities due to income-sharing by farm families not a constraining factor in growth.

Dualism is one example of the application of welfare economics or the theory of market failure to help understand economic development. Another application of this point of view gave rise to the idea of balanced growth, strongly advocated by Ragnar Nurkse and given analytic expression by Tibor Scitovsky. Industry A will not develop in the absence of assured demand from industry B or from consumers. But industry B or the consumer sector in turn will only develop with assured or expected demand which, in turn, will come from the development of Industry A.

Parenthetically, the form in which the argument has just been stated implies that the balanced growth is entirely domestic. It was presented a time when there was pessimism about the prospects for exports by developing countries. Obviously, if the source of some demands is partly

external to the country, the chicken-and-egg problem is not as acute. But I want to avoid the foreign trade aspects of development in this address, not because they are not important but because they would take me too far afield.

How are the externalities suggested by the balanced growth argument consistent with the usual theorems about the optimality of competitive equilibrium? Monopoly is not presupposed nor are there any technological externalities of the type of pollution in the argument. Technically, the market failure is the non-existence of forward markets for the products of each industry. If steel is needed for the production of automobiles, while no one will invest in automobile production in the absence of a domestic source of steel, all that is needed to get started are markets today for delivery of automobiles and steel tomorrow. The steel producers will invest today on the basis of simultaneous sale of their product for delivery in the future, while the automobile manufacturers will put up their plants with the security of steel deliveries when the plants are ready to produce.

In the absence of forward markets, expectations perform much the same role. Indeed, trial and error by individual producers will get the economy to a balanced resource allocation without intervention, if constant returns to scale prevail. If a steel producer starts producing on a small scale, risking little, automobile producers will enter on a small scale, encouraging further entry on both sides. There is a loss of welfare, of course, compared with a development fully coordinated by a complete set of intertemporal markets or by an ideal planner, but the welfare loss will not persist and cannot be an explanation for large long-run differentials in development.

As Hollis Chenery showed, the market will fail to coordinate if there are elements of increasing returns. More generally, the existence of increasing returns is recognised as one of the basic difficulties in both the viability of competitive equilibrium and the use of the price system to achieve a prescribed efficient allocation. The doctrines of imperfect competition developed by Joan Robinson and Edward Chamberlin in the 1930's were developed to show how an economy with many competing firms could exist in the presence of increasing returns.

From the perspective of development, the concept of increasing returns seems to offer an endogenous explanation for economic growth. Indeed, as we all know, Adam Smith emphasized that growth in productivity was due to division of labour, which is limited by the size of the market. This is a perfectly clear statement of the role of indivisibilities, which is virtually synonymous with increasing returns. The idea dropped out of his classical successors; it has no role in Ricardo or Marx, where increases in productivity are solely the result of capital accumulation. With the diminishing returns and the consequent stationary state envisaged by Ricardo and John Stuart Mill, long-run international differences in income levels arise only from differences in natural resources and in the socially determined minimum wage that governs reproduction. Marshall reintroduced increasing returns into economic theory but had difficulties reconciling

them with perfect competition. In his model, increasing returns were always an externality, but as Knight later argued, this cannot be universally true. Every externality is someone's internality. Allyn Young, in his 1928 presidential address to the Royal Economic Society, stressed increasing returns as a major factor in economic growth.

Though Smith and Young were seeking to explain the remarkable development of Europe and its overseas extensions, the doctrine might be equally capable of explaining the differences between developing and developed countries today. Any simple growth model of the type made familiar to us by Robert Solow shows that, under increasing returns, a country that came in to the growth process will always remain ahead, other things being equal. Of course, this rigidly deterministic model does not reflect the fact that leadership does change as the decline of Great Britain and the rise of Japan attest. But it does give an explanation of persistent international differences due only to the timing of entry on to the path of development.

The presence of increasing returns has other explanatory values. It helps explain why there is so much trade among technologically similar countries. The comparative advantage which drives the trade is derived from scale economies in particular, quite narrowly defined industries. The notion of increasing returns should of course, in many cases, be given a dynamic interpretation, so-called learning by doing.

The proposition that increasing returns are the or even a key element in understanding economic growth has striking policy implications. Since increasing returns inhere in very specific industries, including agriculture and services, one implication is that development should not be broad-based or balanced but rather directed to narrow sectors, to achieve economies of scale in one sector before going on to another. Obviously, such a policy is going to be constrained by the demands and needs of final consumers, unless foreign trade can provide substitutes for the underdeveloped sectors.

The implications for planning are ambiguous. On the one hand, the optimal allocation under increasing returns will not be obtained under free markets. In fact, as I have already remarked, the competitive equilibrium is not even viable, and the outcome will be some kind of imperfect competition. But optimal allocation under increasing returns is difficult. For one thing, the optimization requires the use of integer programming, a procedure intrinsically more complex than linear programming and impossible to carry out for large systems even with the most powerful computers. Even more serious are the data demands. The information needed is widely dispersed among the industries and cannot be effectively communicated. It is for these reasons that decentralized decision-making with some element of monopoly is likely to be more efficient.

In conjunction with the theme of this Congress, it is worth noting that all of the points of view that I have discussed can be used to justify allocative biases for industry as against agriculture or services. Unemployment presumably absorbed in agriculture. This argues for development

of industry, but, as experience confirms, even more strongly for services, including construction. The dual economy hypothesis gives the clearest argument for protecting industry, especially in the broader version in which urban service workers, sometimes not easily distinguished from the unemployed, are considered part of the traditional sector. In the balanced growth model, there is no need for special intervention to enable agriculture to meet demand; it is industry that needs encouragement to get started and therefore be a source of demand. It is usual to associate increasing returns with industry as the late Nicholas Kaldor always held, but from a broader point of view, the argument is not so clear. Increasing returns in a form needed to explain long-term economic development and international income differences must be understood to include induced innovation, that is, innovations not worthwhile developing until the scale of application becomes high enough. Here, agriculture has proved to be a powerful stimulus to innovation.

I have talked in a very general way, as befits the occasion, of the relation of some important early strands of economic development theory to developments in the economic theory of developed countries, especially neoclassical economics but also Keynesian theory. To a great extent, neoclassical economics has developed its own internal critiques, which have enriched our economic understanding and led to applications beyond the original intended scope. It is for others, who know the facts of economic development much better than I do, to assess the role of the theories just discussed in understanding and guiding economic development.