ARTHUR LEWIS’S CONTRIBUTION TO DEVELOPMENT THINKING AND POLICY

by

GUSTAV RANIS
Yale University

Arthur Lewis’s seminal 1954 paper and its emphasis on dualism appeared at a time when neither the work of Keynes and Harrod–Domar nor the later neoclassical production function of Solow seemed relevant for developing countries. As a consequence, his model, rooted in the classical tradition, plus its many extensions, generated an extensive literature at the center of development theory. The approach also encountered increasingly strong criticism, some of the ‘red herring’ variety. Some, spearheaded by neoclassical microeconomists like Rosenzweig, raised serious challenges and focused especially on its labor market assumptions. This paper reviews this landscape and asks what theoretical or policy relevance the Lewis model retains for today’s developing countries.

1 Introduction

As is well known, the rebirth of the sub-discipline of development economics coincided more or less with the early post-Second World War era. It is also relevant to recall that this revival of development theory and policy heavily emphasized the breaking of colonial ties which were associated, somewhat erroneously, with the workings of the market and, consequently, placed major emphasis on the role of the state in the newly independent countries of the third world. Unfortunately, the tool kit available to development economists of the day was also fairly limited. On the one hand, there was the Harrod–Domar (Harrod, 1939; Domar, 1957) model, focusing basically on the steady state properties of the developed economy, with little possibility for alternative technology choice and even less for the role of prices, relying heavily instead on savings-pushed growth competing with population growth. Full employment, market clearance, and perfect competition were assumed. On the other hand, there was the Keynesian (1936) model, focusing on advanced economy cyclical issues. Although, as Albert Hirschman (1982) has pointed out, Keynes deviated from the neoclassical mono-economics, full employment orthodoxy of the day, he focused on the temporary unemployment of both capital and labor in the advanced economy, not the secular underemployment of labor in the developing world. Clearly, savings-oriented one-sector models were in vogue, incorporated in both approaches, accom-
panied by a pronounced elasticity pessimism clouding the prospects of agriculture and of exports as part of the intellectual package.

Arthur Lewis, of course, was an active participant in various dimensions of the then current search for applications of existing theory to the problems of the developing world. Very learned and conscious of economic history as few economists of his day, he relied on real world experience and observation; he was interested in and contributed to development planning, to the developmental role of education, and to the analysis of North–South relations, showed considerable sympathy for the Prebisch–Singer (Prebisch, 1962; Singer, 1950) immiserizing growth approach to international trade, and was, of course, responsible for the comprehensive and definitive *The Theory of Economic Growth* (1955), a major contribution to the early postwar revival of interest in the subject. In this paper, however, we will focus on his signal seminal work, that of ‘Development with Unlimited Supplies of Labor’ (Lewis, 1954), not only because this conference was clearly called to commemorate the fiftieth anniversary of that famous May 1954 paper but also because it is generally accepted as the contribution for which Lewis received the Nobel Prize and with which he revolutionized contemporary thinking on development. In Section 2, we will examine Lewis’s classical roots as well as his less well-known deviations from the classical tradition. In Section 3, we will attempt to outline the impact his thinking has had on development theory, including extensions and criticisms. In Section 4, we will focus on his model’s current theoretical and policy relevance.

2 **Classical Roots and Lewisian Offshoots**

It is usually claimed that Lewis’s simple model is based on classical school foundations, i.e. it contains two sectors, agriculture and non-agriculture, with different, i.e. asymmetric, behavior postulated for each. This basic heritage notion is generally accepted, but, if one looks more closely, one will also see substantial deviations. For example, the classical school actually depicted agriculture as a capitalist sector, with three factors: capitalist farmers renting land from landlords, and hiring free labor. Landlords get their rents via the neoclassical marginal productivity principle, and capitalists then bargain with workers on how the rest is apportioned. Landlords are seen as wastrel consumers. The non-agricultural sector, fed by capitalist profits, is not really modeled but, except in Smith’s (1880) more optimistic view, represents but a temporary deviation from ultimate agricultural stagnation, resulting from population growth squeezing out capitalist profits in the absence of reliable technology change. The classicists, Ricardo (1815) in particular, were clearly looking over their shoulders rather than at the revolutionary changes beginning to take place all around them in both sectors.

All this is quite in contrast to Lewis’s model, which, of course, also depicts a two-sector world, also built on physiocratic as well as classical
antecedents but in which agriculture is now the dominant non-capitalist, or subsistence, sector, with only two factors at play, landlords and workers, and wages set in a bargaining context. Lewis, like Smith, saw the relatively small non-agricultural commercialized sector as potentially dynamic and expanding, but he was much more optimistic concerning the ability to mobilize the hidden rural savings of Nurkse (1953) and Rosenstein-Rodan (1943) in a static context as well as via productivity change in agriculture. Landlords were seen as a saving class as well as potential commercial and industrial entrepreneurs. Non-agricultural growth here is seen as vigorous and sustained and not, as in the classical system, a temporary deviation from basic agricultural pursuits.

Lewis focused on organizational dualism and much less explicitly on product dualism. Indeed, neither Lewis nor the classical school concerned themselves in detail with the analysis of intersectoral relations or the intersectoral terms of trade. Lewis’s main focus was on the reallocation of labor until the turning point is reached, i.e. the time when labor reallocation has outstripped population growth long enough for dualism to atrophy and the economy to become fully commercialized. The fact that the terms of trade are a crucial determinant of intersectoral labor market, financial market and commodity market clearance is not something he very much concerned himself with.

On the other hand, Lewis really moved beyond the classical school in a number of important dimensions. First, he was interested in transition growth from a dualistic to a one-sector, modern economic growth world in the Kuznets (1971) tradition, from organizational dualism to organizational homogeneity, i.e. he saw the development problem as focusing on a change in the basic rules of operation of an economic system. Second, he believed in the power of technology operating in both sectors, although he did not model it explicitly. Third, he rejected the neo-Malthusian (1815) heritage of the classical school; and finally, although not an explicit part of his basic model, he pointed out that food shortage problems could be overcome by imports in the open economy.

3 Lewis’s Impact on Development Theory

The basic labor surplus model was, of course, very simple, elegant and to the point, a true reflection of the man. Arthur Lewis never favored formal theorizing or complicated diagrams; he did not feel the need to present well-specified mathematical models. Where he excelled was in the strength of his intuition and his sense of history. He knew how to get to the heart of the matter and, in the process, succeeded in making economic development respectable in a number of ways. For one, he was one of those early birds who helped move this neglected sub-field of development economics away
from the neglect of prices and the lack of faith in the potential for agricultural productivity change and exports. He did not share the commonly held belief in an all-powerful state which was expected not only to create the pre-conditions for development but also to organize most of the required directly productive activities. Lewis clearly saw the overarching need for private actors to complement government planners.

Lewis moreover rejected the neoclassical assumptions of full employment, market clearance and perfect competition, even as he saw it as a distant goal, along with Ken Arrow (1962). He explicitly recognized that not only owner-operated agriculture but also the urban informal sector, lacking cooperating capital instead of land, was characterized by a system of bargaining rather than competitive wages. Most importantly, he opened the door to an extensive literature focusing on both extensions and criticisms of what was for a long time considered the basic model of development. Moreover, as Minami (1973), Ohkawa (1972) and Fei and Ranis (1964), among others, have pointed out, Lewis really contributed in a major way to transition growth theory, to the notion of development phases and sub-phases, en route to modern economic growth.

Turning first to extensions, Lewis, as has already been mentioned, focused mainly on organizational dualism, on intersectoral labor markets explicitly and on intersectoral financial markets implicitly; he had relatively little to say about intersectoral commodity markets and the intersectoral terms of trade, which was left to Fei and Ranis (1964), among others, to explore. In fact, the importance of balanced growth between the two sectors, while implicit in his reasoning, could really only be pinned down by superimposing product dualism on his organic dualism so that food shortage could lead to a rise in the real agricultural and, consequently, the unskilled industrial real wage, before the Lewis turning point, signaling the exhaustion of labor surplus, could be reached.

Other extensions of the basic Lewis model can be found in the Harris and Todaro (1970) and Fields (1975) contributions. Harris and Todaro’s main innovation was to introduce the notion that intersectoral labor reallocation is affected not only by the intersectoral wage gap but also by the probability of obtaining a formal sector job. They accepted the idea of institutional interventions in determining the level of the non-agricultural urban unskilled real wage, arising from union, minimum wage and government wage setting, consistent with Lewis’s intersectoral ‘hill’ concept, but they insisted on a competitive neoclassical agricultural wage, in contrast to Lewis. Fields (1975), closer to Lewis’s basic model, pointed out that there were three choices for migrants: a formal sector job or open urban unemployment, plus a third possibility, a job in the urban informal sector, which Lewis had already pointed to. Just as in agriculture, he stated that very few urban residents can afford to be openly unemployed and rely on usually non-existent unemployment.
insurance. Instead, just as in agriculture, they fall back on family sharing, while working at low levels of productivity, i.e. they are the urban underemployed. Among other extensions we would count the work by Ranis and Stewart (1999), which differentiates among two urban informal sub-sectors, a V-goods sub-sector which is dynamic and tied by subcontract to the urban formal sector, and an informal sponge sub-sector which was the focus of both Lewis and Fields.

Lewis’s model also had implications for income distribution, very much in line with Kuznets’s (1955) early contribution to the subject. Kuznets’s structural analysis, as the economy moves from agriculture to manufacturing to services, implicitly also adopted a dualistic model. His reasons for anticipating an initial worsening of income distribution was that, as labor shifts from an equally distributed agricultural to a less equally distributed non-agricultural sector, this leads to a worsening of the overall distribution until wages rise in a one-sector world. This makes it very much akin to Arthur Lewis’s view, which also has distribution likely to be worsening as long as wages are depressed, i.e. before the Lewis turning point is reached. Neither Lewis nor Kuznets can be said to have taken into account the possibility that the employment effects of low wages during the early reallocation process can, in fact, lead to an increase in the wage bill and a functional distribution favoring labor, which can lead to an improvement in the family distribution of income—see the experience of Taiwan, for example. It is nevertheless clear that the Lewis model had substantial influence on subsequent work on the relationship between growth and equity.

Finally, last but not least, we should note that the Lewis model has also been applied to labor movements across countries, along with movements between two sectors in the closed economy. Kindleberger (1967), for example, used the Lewis model to describe the migration of surplus labor from the Maghreb countries of Northern Africa and Turkey to Europe during the postwar boom of the European Community. Indeed, the flow of surplus labor across borders, along with the impact of remittances, remains one of the more controversial issues in development theory and policy to this day.

Turning to some of the critiques of the Lewis model over the last 50 years, it must be admitted that, while the model has been widely praised, it has also come under severe attack and is today less frequently quoted, certainly in Anglo-Saxon mainstream economics. This is largely due to the shift in the sub-discipline to an emphasis, on the one hand, on macroeconomic cross-section analyses such as that of Barro (1991) and, on the other, on micro-econometric analysis with firm adherence to the neoclassical mono-economics paradigm, accompanied by the steadfast rejection of any bargaining theory of wages. Some of these criticisms, however, are due to misunderstandings caused, not so much by Lewis himself, who was rather clear and cautious, but by Lewisians who were sometimes less careful and consequently opened up the model to what might justifiably be called ‘red herring’ attacks.
The most serious of these is probably the notion that ‘labor surplus’ was interpreted as zero marginal productivity of agricultural labor, a highly unlikely event, statistically or conceptually, and one which was subjected to vigorous attack by Ted Schultz (1964) who introduced evidence from India to show that the withdrawal of a large portion of the agricultural population did not lead to a decline in agricultural output. This claim was repudiated by Sen (1967) who pointed out that as people leave agriculture those who remain may work harder. A broader statement would be consistent with the view that any withdrawal of labor from agriculture is likely to be accompanied by a reorganization of production by those who are left behind, i.e. technology change. Lewis thought of labor surplus in terms of human beings rather than man-hours, and his labor surplus was really defined in terms of an excess supply of labor at the going wage, a concept consistent with the Fei–Ranis emphasis on wages in agriculture exceeding the marginal product, which might be quite low, even if not zero.

Closely related is the criticism of Ootsuka (2001), among others, who states that he has, ‘never encountered institutionally determined rigid wage rates in agrarian communities’. If agricultural wages are indeed determined by the sharing of income, i.e. related to the average agricultural product, we would expect to see a gradually rising real wage as the bargaining solution takes into account rising levels of that average product. Consequently, what you get is a step function, with gently rising real wages, but still not keeping up with productivity which is rising faster. The basic point is that it is the sharing rule which matters, not the level of a wage which is likely to vary over time. The dual economy model moreover assumes that agricultural wages are related to, but not necessarily equal to, the average product of agricultural workers, with the head of the household or commune, or whoever else commands the agricultural surplus, retaining a certain portion for his or her reinvestment purposes. Consequently, Arthur Lewis’s supply curve in the real world is gently rising over time and therefore not horizontal, which it is only at a given point in time. Over longer periods, we see a step function made up of annual unlimited supply of labor segments, econometrically difficult to distinguish from a gently rising supply curve.

Another criticism which has been leveled against the Lewis model is that the classical assumption that all profits are saved and all wages are consumed has been retained. This assumption is quite unnecessary to the basic model and is simply stipulated by Lewis and by Lewis’s followers as a convenient simplification. Similarly, the notion that there is no physical capital accumulation in agriculture, with all the investment going into non-agriculture, is neither essential nor empirically correct, again constituting only a simplification.

The most serious objection to the Lewis model, of course, is that contemporary development economists, working inside the neoclassical paradigm, cannot accept the notion of an exogenous unskilled agricultural real
wage instead of one that is determined endogenously by the interaction of demand and supply. When they do accept the notion of institutional interventions, it is not a wage set by bargaining but one having to do with insurance, over either space or time. This is the crux of the contemporary critique of the Lewis model, i.e. the rejection of an exogenous bargaining wage or consumption share, exceeding the marginal product of labor at any point in time. Lewis, Fei–Ranis, and other followers accept the notion that in the particular setting of a heavily populated agricultural sector, with an extended family or other communal institutional arrangements, the unfavorable ratio of people to collaborating factors is part of the initial condition which makes it impossible for decision makers to simply fire low productivity group members in order to reach a neoclassical equilibrium or to somehow refuse to share the group’s income with them. Unfortunately, there is no acceptable model which yields a uniquely determined bargaining wage. But this also holds in the advanced economy union/management bargaining context, yielding ultimately a marginal productivity solution, even if the time lag in that case is, of course, substantially shorter. In the absence of a neat theory to determine the level of a bargaining wage which may hold for some decades, the Lewis model has been consistently praised, but also increasingly rejected.

Yet the basic sharing assumption gets support from anthropologists like Geertz (1963) and Scott (1976). The economist Ishikawa (1975) endorses the concept of a minimum subsistence level of existence, which is but one version of the institutional real wage and is defined by him as a ‘community principle of employment and income distribution’ which promises all families an income not less than a minimum subsistence level. Hayami and Kikuchi (1982), while basically neoclassical in approach, find that, in Indonesia, ‘wages do not adjust on the basis of labor’s marginal product but according to the subsistence requirements of the time and social conventions’. Only over time, perfectly consistent with Lewis, is there a tendency to adjust by including weeding duties without a complementary rise in the wage. Osmani (1991) presents a model of downward rigidity of the sharing rule as insisted on by the workers; and current work in what is called behavioral economics may prove to be of help in developing a theoretical structure to rationalize cross worker subsidization in the absence of assured reciprocity, especially if some members of the group are likely to be leaving agriculture over time.

Finally, and perhaps most importantly, there is historical evidence, including for England between 1780 and 1840, for Japan between 1870 and 1920, for Taiwan between 1950 and 1970, of labor-abundant agriculture exhibiting large increases in average agricultural labor productivity while the agricultural and non-agricultural real wages rose only gently, i.e. lagging substantially behind, until the commercialization or Lewis turning point is reached. As Sen (1966) has pointed out, even a horizontal supply curve of labor can be made consistent with neoclassical explanations, but you have to
work hard at it. These facts are fully consistent with an institutional wage which is gently rising as a result of the step function previously described, until the turning point is reached and wages begin to rise steeply in concert with rising marginal productivity. Before that point is reached, a rising gap between agricultural productivity and wage levels is certainly not consistent with neoclassical assumptions about labor market clearance.

Mark Rosenzweig (1988), among others, has presented microeconometric evidence of steeply rising labor supply curves in a cross-section of heavily populated agricultural economies, such as India, and claims that this has put the final nail in the coffin of the classical dualistic model. But the Lewis supply curve is merely a facet of the operational interaction between two sectors of a dualistic economy. The labor absorption path is derived from the time path of the industrial wage and really constitutes the locus of various combinations of wage rates and the labor absorbed, associated with different levels of the industrial capital stock and the level of technology. It is based on the time path of the industrial real wage, with a gently rising and then a steeply rising portion, which can be defended empirically and theoretically at a macro level and is actually quite irrelevant to the focus of timeless microscopic household studies in the Becker tradition.

4 Current Theoretical and Policy Relevance of the Lewis Model

It should be of some interest to note that the Lewis model and its many offspring continue to be viewed as relevant in the South and considered a valuable guide to policy in places like China, India, Bangladesh, Central America and even some parts of sub-Saharan Africa, i.e. wherever heavy population pressure on scarce cultivable land remains a feature of the landscape. Most Northern development economists, on the other hand, are today focusing either on aggregate cross-section models to determine the sources of economic growth in the Barro (1991) tradition or, at the micro level, on the econometric modeling of household behavior, with very little interaction between the two approaches. In the South, dualism still holds the attention of both theoretical and empirical observers. Bourguignon and Morrison (1995) still see the persistence of economic dualism as a powerful explanatory factor underlying cross-country differences in inequality in the Lewis and Kuznets tradition, explicitly or implicitly embracing the dualistic model, with wages kept relatively low and savings rates rising as long as there is a labor surplus, followed by the eventual improvement of equity with the upswing of real wages. This yielded, in Kuznets’s view, the likelihood of the famous inverse U-shaped pattern, depicting the relationship between growth and the distribution of income over time. Work by Fei et al. (1979), Fields (2001) and others has shown that indeed no inevitability attaches to the Kuznets curve. But it is also clear that the nature of the growth pattern itself
needs to be viewed in an expressly dualistic context to determine the relationship between the functional and the family distribution of income over time, all of which differs markedly in the period before and after the Lewis turning point.

The relationship between growth and equity in the Lewis tradition, of course, also spills over into an analysis of technology choice, technology change and the relationship between growth and poverty, currently very much on the front burner of both theorists and policymakers. Low real unskilled wages in agriculture and industry and the expectation of more to follow, of course, favor labor intensive technology choices statically and labor using technology change dynamically. This fact, and the reversal in these dimensions once the system enters the one-sector neoclassical world after the Lewis turning point, has been documented. The asymmetry between sectors and the interest in the contrasting pre- and post-turning point behavior of the whole system clearly supports the dualism model theoretically as well as being helpful to policymakers. For example, the issue of the intersectoral terms of trade and the importance of balanced growth policies, which need to be more or less maintained before the turning point in order to avoid food shortages, continues to be of importance in the contemporary development context, even in the open economy. Food imports do not solve the problem of a failure to mobilize the agricultural sector on behalf of a successful development effort. Indeed, they may contribute to the problem.

The assumption of the persistence of an abundant supply of labor over some historical time period also affects the open economy dimensions of development in another respect. According to Lewis, productivity changes will accrue to the importing or advanced country, leading to another version of immiserizing growth. This is one area in which Lewis’s adherence to Prebisch–Singer probably did not sufficiently take into account the difference between labor intensive industrial and agricultural exports—although he properly emphasized the growing potential for inter-LDC trade. All in all, Lewis rightly saw technology, not trade, as the more dependable engine of growth.

Surprisingly, the Lewis model of dualism also has some relevance to contemporary mainstream development models at the micro level. For example, the ‘informal insurance mechanism’ of Townsend (1994) by which farmers smooth consumption by insuring each other across space is not radically different from the aforementioned ‘moral peasant’ of Scott (1976) who is concerned with supporting others over time as well as space. Whether all this can be forced into a comforting neoclassical model or comes close to institutional altruism remains a point of contention, and one would hope that the current emphasis on the new institutional economics could potentially be an ally of the revival of the concept of dualism as an important guide to development theory and policy. In Townsend’s world, income is reallocated *ex post*, i.e. after neoclassical distribution rules have been observed,
while in the Lewis world, income is divided ex ante among members of the extended family or wider community. The policy implications for achieving a successful transition to modern economic growth probably do not differ fundamentally depending on which of the concepts is deployed. But what remains relevant is which model fits better the basic empirical reality in successful labor-abundant countries: which is better suited to analyzing agricultural neglect in failure cases; which provides a better explanation of the marked early rise in the system’s savings rate; which is more capable of explaining discontinuities in income distribution and technical choice and the direction of technology change—a model that assumes full employment and smooth neoclassical equilibrium everywhere or one that recognizes initial underemployment and disequilibrium en route to a one-sector modern economic growth epoch.

Lewis was basically a macro-economist, deeply immersed in economic history and the history of thought, both neglected subjects today. He always chose a general equilibrium approach, not only with respect to working within a domestic two-sector world but also with respect to the relationship of the typical developing country to the world economy, as indicated by his Wicksell and Janeway lectures (Lewis, 1969, 1977). His notion of dualism, especially that focused on the labor market dimension, rural and urban, continues to offer a theoretically valid, empirically relevant and practically useful framework for dealing with some fundamental real world issues of development.

References


