Chapter 18

IS DUALISM WORTH REVISITING?

Gustav Ranis
Yale University, gustav.ranis@yale.edu

1. INTRODUCTION

In the 1950’s and 1960’s, a neglected sub-field of economics was rediscovered: development economics, concerned with describing the economies of the world’s poor countries, which even then accounted for more than half of the world’s population. The economic problems facing these countries—many of them newly independent—were clearly acute and their development was seen as an urgent priority for both analysis and policy action.

Available economic models, however, seemed to offer only limited insights into the practical problems facing these countries. These were not modern industrial economies; most people worked not in factories but in subsistence agriculture or in urban informal jobs—from shoe shining to small-scale peddling to distributive trades and artisanal crafts. The dominant one-sector macro models of the day, from Keynesian to Harrod-Domar to Solow, seemed to have limited relevance for societies not primarily concerned with business cycle or steady state properties. Most contemporary growth models were seen as academic abstractions with little policy relevance. And the dominant assumptions of neoclassical micro theory—full employment, market clearing and perfect competition—seemed to have little relevance for the segmented commodity, labor, and credit markets of poor countries.

Against this backdrop, the concept of dualism attracted considerable attention. Sociological dualism, associated with the name of Boeke, emphasized differences between Western economic and non-Western cultures and objectives. Technological dualism, emphasized by Higgins,
technology change, either generated within agriculture or via modern inputs from non-agriculture, agricultural stagnation and thus the drying up of the needed agricultural surplus represented the dominant long-term outlook. Whether technology change and the exploration of economies of scale within the industrial sector, reflecting Smith’s optimism, would be sufficiently strong to provide enough industrial profits to rescue the situation remained controversial.

It was, of course, Arthur Lewis who in his famous 1954 article built on some of the main ingredients of this classical tradition, leading him to emphasize dualism in labor markets, i.e., a competitive wage in non-agriculture but tied to a wage in excess of a very low, if not zero, marginal product in agriculture. Lewis, moreover, found himself allied with Smith, seeing the relatively small non-agricultural or commercialized sector as dynamic and expanding, fed by the mobilization of the “hidden rural savings” of Nurkse and Rosenstein-Rodan, enabling the reallocation of workers into higher productivity activities, while wage levels were kept relatively low, at a modest “hill” over the agricultural wage. This reallocation process would continue until all the “surplus labor,” i.e., all those whose remuneration exceeded their marginal product, had been reallocated, marking a turning point at which dualism atrophies and the economy becomes fully commercialized or neo-classical.

In Section II, we will trace the further development of the dual economy model in the Lewis tradition. In Section III we will examine the criticisms which have been leveled against it by the neo-classical school, differentiating between what may be referred to as popular “red herring” attacks and more central critiques. Section IV will inquire into the current normative usefulness of the concept for both analytical and policy purposes. Section V concludes.

2. THE MODERN DIMENSIONS OF DUALISM

We will concentrate here on dualism in the labor market and, for reasons of convenience, restrict ourselves to the closed economy—relevant to all but very small economies. Dualism, of course, matters only when an economy’s agricultural sector is initially relatively large and represented substantially by extended family or collective farm cultivation, often labeled “subsistence agriculture,” in contrast to commercialized or plantation agriculture.

The critical organizational characteristic of this sector is that, at a given technology, the man/land ratio is such that marginal labor productivity is very low (if not zero or negative), while key decision makers such as heads
of families, village elders or commune leaders feel obliged to share output with all members of these groups, even if such shares exceed the marginal product. In other words, income is shared, or an institutional wage is determined, based on bargaining rather than neo-classical principles. Moreover, the gap between this institutional wage in non-commercialized agriculture and the commercialized non-agricultural wage is likely to be in excess of the modest 50% “hill” asserted by Lewis and is materially affected by such institutional urban formal sector interventions as minimum wage legislation, union pressure and government wage setting. Extending the Lewis model further, there are likely to exist similar configurations in the so-called urban informal sector, with family enterprises lacking sufficient cooperating capital and forced to pursue service sector and distributive trade activities which yield low marginal labor productivities but are again characterized by a similar pooling of income in an extended family context. In both cases everyone “in the same kitchen” is fed and average, rather than marginal, product is relevant to the size of the shares.

In addition to the organizational dimensions of dualism, emphasized by Lewis, there is a product dimension, focused on the exchange between food produced by the peasant agricultural sector and the non-agricultural goods produced in the urban (and/or rural) commercialized non-agricultural sector. The interaction between these two sectors, extending beyond the inter-sectoral labor market into the inter-sectoral commodity and financial markets, was fully analyzed by Fei and Ranis. The key point here is that agricultural and non-agricultural products cannot readily be substituted for each other; in the closed economy food-producing agriculture becomes a necessary condition for industry, while the converse does not hold. Consequently, if agriculture lags behind non-agriculture during the labor reallocation process the deterioration of the non-agricultural sector’s terms of trade may well cause a rise in the agricultural wage, as well as in the related unskilled non-agricultural real wage, and a consequent retardation if not collapse of the labor reallocation process long before the labor surplus has been eliminated. This “turning point,” in contrast to Lewis’, signals the relative inability of the system to follow a “balanced growth” path according to which agricultural productivity keeps pace with advances in non-agriculture and inter-sectoral commodity as well as financial and labor markets can clear at given terms of trade.

Other modifications of the basic Lewis model can be found in Harris-Todaro and Fields. While Harris-Todaro’s main innovation was to introduce the notion that labor reallocation is affected not only by the inter-sectoral wage gap but also by the probability of obtaining a formal sector non-agricultural job, they accepted the notion of an institutional impact on the level of the non-agricultural urban wage but insisted on an equally competitive, i.e., neo-classical, agricultural wage. Fields, however, pointed out that the two choices for migrants offered by Harris-Todaro, a formal sector job or open urban unemployment, needed to be amended by introducing the urban informal sector. Indeed, just as in agriculture, very few urban residents can afford to be openly unemployed and rely on non-existent unemployment insurance. Instead, they fall back on the family, while working at very low levels of productivity, i.e., they are the urban underemployed.

Finally, Ranis and Stewart further amended the basic dualistic model by suggesting the need to trace the interactions among four sectors once we further disaggregate the urban informal sector between a non-traditional dynamic sub-sector, with sub-contracting ties to the urban formal sector, and a traditional static sub-sector, serving as a low productivity sponge. The very meaning of “dualism” of course becomes somewhat tenuous as we move away from the basic two-sector Lewis model but, even though inter-sectoral relations become increasingly mind-blowing to trace, the basic asymmetries in labor market behavior remain critical for both analytical and policy purposes.

2.1 Neo-Classical Critiques and Responses

Dualism has been subjected too much criticism and attack over the past several decades. It has virtually disappeared from contemporary development discourse in the OECD countries, except via textbooks including intellectual history. Some of this may be termed not so much malign neglect as a consequence of the development sub-discipline, along with the rest of economics, moving away from grand macro-theorizing to a micro-econometric focus. But much of the attack has also focused quite specifically on the theory of dualism because of its “unacceptable” assumptions about labor market behavior. In spite of its apparent real world empirical relevance—which more below—the critics reject out of hand any bargaining outcome which cannot be modeled precisely within a neo-classical framework.

Some of the specific critiques on record may be viewed as “red herrings” and can be readily responded to; others are more profound and require more careful consideration. Perhaps the leading attack in the “red herring” category was occasioned by the unfortunate choice of the “labor surplus” term deployed by Nurksa, Lewis, Fei-Ranis and others. This was widely interpreted as implying a zero marginal product in agriculture and led to the famous T.W. Schulz/Sen exchange on whether or not a reallocation (or in this case demise) of part of the agricultural labor force could be expected to leave agricultural output unaffected. This misperception arose due to
incautious asides in Lewis' 1954 paper and the resort to mathematical convenience in Fei/Ranis' 1961 and 1964 contributions. The basic point is that the marginal product is low, and sufficiently low to fall below the bargaining wage or income share. As Lewis put it in his 1972 retrospective piece\textsuperscript{16} "whether marginal productivity is zero or negligible is not at the core of fundamental importance to our analysis...this has led to an irrelevant and intemperate controversy." But the controversy has persisted. Otsuka in his review of Fei/Ranis' 1997 book\textsuperscript{17} considers the zero marginal product notion as deeply embedded in the dual economy model but acknowledges that such a stark assumption is both empirically unlikely and theoretically unnecessary. What is necessary is that, during any short period of time, there exists an excess supply of labor at the going wage.

Does this mean that the reallocation of labor necessarily causes a food shortage, or the arrival of the Fei-Ranis "shortage point?" Again, as Fei-Ranis took pains to point out, a withdrawal of labor is highly likely to lead to a simultaneous reorganization of agricultural production, in effect an upward shift in agricultural labor productivity, permitting the maintenance, if not even an increase, of the agricultural surplus available for transfer to non-agriculture. Thus, the inter-sectoral terms of trade need not deteriorate against industry—unless, of course, there is a marked relative neglect of agriculture in the "balanced growth" context.

A second critique, again in the "red herring" category, dealt with the level of the exogenously given institutional wage. The neo-classical school, of course, fundamentally rejects the notion of an institutional or bargaining wage since it cannot be deduced from basic principles. But the fact that wages are, in fact, observed as not constant over time can be easily disposed of. Otsuka,\textsuperscript{16} for example, claims he has "never encountered institutionally determined rigid wage rates in agrarian communities". This ignores the fact that it is the sharing rule, not the level of a wage, which may well vary over time, which is at stake. The dual economy assumption is that agricultural wages are related to, but not necessarily equal to, the average product of agricultural workers, since the head of the household, or whoever else commands the agricultural surplus, is bound to retain a portion for her own reinvestment purposes. As the average product rises with technical change, the agricultural bargaining wage is also likely to change. Thus, over time, we are likely to see a gently rising, not a horizontal, Lewis-type supply curve of labor. Over short periods of time it is horizontal; but what we have over longer periods is a step function made up of annual unlimited supply of labor segments, econometrically undistinguishable from a gently rising supply curve.

Let me cite, in passing, a few other criticisms of some assumptions of the dualism model, all of which have some validity but none of which are critical one way or the other. They include the adoption of the classical assumption that all wages are consumed and all profits are saved; that the system is savings-pushed, i.e., that Say's Law holds; and that all investment funds are allocated to the commercialized or non-agricultural sectors. There would be absolutely no problem in modifying any of these simplifying assumptions without any damage to the basic dualistic model.

Let us then turn to the crux of the critique, the rejection of a bargaining wage or consumption share exceeding the marginal product of labor at any point in time. I assume it is not difficult to see that in the kind of setting under discussion, i.e., extended family or other communal institutional arrangements, the unfavorable ratio of people to cooperating factors as part of the initial condition is not something under the control of decision makers, and that such decision makers cannot simply fire low productivity members to reach a neo-classical equilibrium or refuse to share much of the group's income with them. Fafchamps\textsuperscript{18} provides an overview of the principles underlying the resulting "solidarity network" among peasants, as depicted in the anthropological evidence of Geertz\textsuperscript{19} and Scott.\textsuperscript{20} Ishikawa,\textsuperscript{21} an astute long-time observer of Asian peasants, endorses the concept of a "minimum subsistence level of existence" (MSL), one version of the institutional real wage. His work indicates the prevalence of a "community principle of employment and income distribution which promises all families...an income not less than MSL." Hayami and Kikuchi\textsuperscript{22} 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 is there a tendency to adjust but even then it does not necessarily occur by altering wages to equal the marginal product, which could reduce the wage below subsistence. Instead, in Java harvest contracts began to include weeding duties without a complementary rise in the wage rate, thereby not threatening the MSL but moving institutionally towards equilibrium. Osmani\textsuperscript{23} presents a model of downward rigidity of the sharing rule insisted on by the workers themselves. Current work in what is called behavioral economics may also prove to be of help in developing a theoretical structure to rationalize cross-worker subsidization in the absence of assured reciprocity—especially as some members of the group are likely to be leaving agriculture over time.

Finally, we have ample historical evidence, e.g. for England from 1780 to 1840,\textsuperscript{24} for Japan from 1870 to 1920,\textsuperscript{25} for Taiwan from 1950 to 1970,\textsuperscript{26} of labor abundant agriculture witnessing hefty increases in average agricultural labor productivity, while the agricultural wage rises only gently, i.e., lags substantially behind, until the commercialization or Lewis turning point is reached. As Sen\textsuperscript{27} has pointed out, even a horizontal supply curve of labor can be made consistent with a neo-classical explanation; but you
have to work hard to make the pre-conceived theory fit the facts. And these facts are also fully consistent with an institutional wage which is gently rising as a result of the step function process previously described until, as in the country cases cited above, the turning point is reached and wages begin to rise steeply in concert with rising marginal productivity. But before that point is reached, a rising gap between agricultural productivity and wage levels is certainly not consistent with neo-classical assumptions about market clearance.

Rosenzweig and others have presented micro-econometric evidence of steeply rising labor supply curves in a cross-section in heavily populated agricultural settings such as India and claim that this puts the final nail into the coffin of the classical dualistic model. However, this is an expressly static relationship and does not address a developing economy’s dynamic transition process. Moreover, as we have shown elsewhere we would expect individual family labor responses to be quite inelastic at any point in time. Faced with hypothetical wage changes, hard-working, if not highly productive, agricultural family workers are not likely to have much room for trading off leisure for additional work. Rosenzweig’s findings are inherently reasonable but they address a different issue. He is concerned with the cross-sectional labor/leisure decisions across agricultural households, while dualistic models are interested in the conditions governing inter-sectoral labor reallocation in a time series context.

3. DOES THE DUAL ECONOMY MODEL STILL SERVE A USEFUL PURPOSE?

Even if it were accepted that dualism is a useful construct for understanding the historical development experience of a number of countries, including England, Japan, and Taiwan, among others, does it have any empirical relevance for today’s world; and, even more importantly, how does it relate to contemporary theorizing in the “new growth theory” or “new institutional economics” traditions?

We would contend that China, India, Bangladesh, as well as much of Central America and some portions of South America, comprising a majority of the people on earth, still meet the initial conditions for dualism relevance, i.e., a substantial food producing agricultural sector marked by heavy population pressure on scarce land, complemented by large urban informal sectors. And, while Sub-Saharan Africa was once described as “land surplus,” there is increasing evidence, from high fertility rates, changing cultivation practices and a general reduction of fallow periods, that this pivotal region is also moving in the same direction. In these settings the issue addressed with the help of the dualism model, i.e., how to mobilize an agricultural surplus by reallocating an underemployed labor force into efficient non-agricultural pursuits, remains at the top of the development agenda.

The dualistic theory nexus, moreover, remains useful for a number of analytical reasons, including concerning the relationship between growth and the distribution of income, for the determination of the domestic inter-sectoral terms of trade, as well as for the choice of technology and of the direction of technology change. The effort to understand the relationship between growth and the distribution of income was first brought to our attention by Simon Kuznets in 1955. While Kuznets emphasized the structural change an economy experiences as it shifts from initially dominant A (agriculture) to M (manufacturing) and S (services) sectors over time, his explanation of the famous inverted U-shaped pattern between growth and distribution embraced the dualistic model. The basic cause for the initial worsening of distribution was the reallocation of workers from a more equally distributed agricultural sector to a less equally distributed non-agricultural sector—with wages kept relatively low and savings rates rising—while the eventual improvement of equity is related to the upswing of real wages as full employment is reached everywhere. Bourguignon and Morrison see “the persistence of economic dualism as a powerful explanatory factor of cross-country differences in inequality.” While work by Fei, Ranis, and Ku, Fields and others has shown that no inevitability should be attached to the suggested inverse U-shaped path, it is clear that the nature of the growth pattern viewed in an expressly dualistic context determines its relationship to equity over time, differing markedly between the period before and after the commercialization point.

This concern about the relationship between growth and equity eventually spilled over into a focus on what is happening to poverty levels. The notion of only gently rising real wages in both agriculture and non-agriculture during the pre-commercialization epoch affects not only labor, commodity and financial exchanges between the sectors but also has a definite impact on technology, favoring labor-intensive technology choices stastically and labor-using technology choices dynamically. The reversals in these dimensions once a system enters the one-sector neo-classical world has also been documented. In all these instances the acceptance of initial asymmetry between sectors rather than the neo-classical assumptions of full employment and smooth homogeneity can prove helpful to open-minded analysts.

The pattern of the inter-sectoral terms of trade provides an indication of whether or not the dual economy has managed to maintain “balanced growth” and avoided the relative agricultural neglect much discussed in the
literature. Finally, we should also note the relevance of dualism for contemporary mainstream development models. The "informal insurance" mechanism of Townsend\textsuperscript{xxvi} for example, by which farmers smooth consumption by insuring each other across space is not radically different from the aforementioned "moral peasant" of Scott\textsuperscript{xxvii} (1976) who is concerned with supporting others over time as well as space. Whether this can all still be forced into a comforting neo-classical bottle or approaches institutionalized altruism, of course, remains a point of contention. In the former, income is allocated \textit{ex post} after neo-classical distribution rules are observed, whereas, in the latter, income is divided \textit{ex ante} among the members of the extended family or wider community. It is unclear whether the policy implications for achieving a successful transition to modern growth differ widely depending on which concept is deployed. But what remains relevant is which model fits better the basic empirical reality of wage behavior relative to agricultural productivity changes in successful countries; which is better suited to analyze agricultural neglect in failure cases; which provides a better explanation of the marked rise in a system's savings rate; which is more capable of explaining discontinuities in income distribution and technical choice—one that assumes full employment and smooth neo-classical equilibrium everywhere or one that recognizes initial underemployment and disequilibrium in the system en route to a one sector modern growth epoch.

4. CONCLUSION

The interaction between the agricultural and non-agricultural sectors resides at the heart of early stage development in many developing countries. We have in this paper argued that dualism, especially focused on its labor market dimension, continues to offer a theoretically valid, empirically relevant, and practically useful framework for dealing with this issue.

After briefly tracing the intellectual history of dualism, we cited and responded to various neo-classical critiques, distinguishing between those which can be characterized as of the "red herring" or straw man variety and those which are more fundamental and thus need to be seriously addressed.

The key assumption at issue is the acceptance or rejection of an institutional or bargaining real wage or income share as an imposed initial condition for an agricultural sector composed largely of owner/cultivators and, given very high man/land ratios, resulting in low levels of marginal productivity. A wage above this marginal product agreed to by the family, the community or the commune, whatever the organizational configuration—and whether in agriculture or in informal urban sector activities—is likely to be related but not equal to the average product—and that is difficult for adherents to the dominant neo-classical school to swallow. We have tried to respond to various criticisms, including the interpretation of surplus labor as zero marginal product labor, instead of simply indicating the existence of disguised unemployment, i.e., income or consumption shares exceeding the marginal product. We have characterized the supply curve of unskilled labor in agriculture and the informal urban sector as a step function composed of horizontal portions, each indicating the "unlimited" availability of labor for commercialized sector absorption over any short period of time. Concerning the more serious challenge to the dualism model, the finding of an inelastic supply curve of agricultural labor, our response was two-fold: that we see no inherent conflict between cross-sectional micro-econometric findings in the neo-classical tradition and dualism's effort to trace the dynamic time series of the inter-sectoral reallocation process at the macro level over time; second, that we find it not surprising that individual agricultural workers, even if underemployed due to the lack of cooperating factors, work long hours and have very little leisure left to surrender in response to a higher wage.

We, finally, inquired into the usefulness of the dualism concept for explaining both historical and prospective country development experience, as well as its relevance in general to contemporary economic modeling efforts. We found that the basic dualism model well fits the historical experience of such countries as England, Japan, and Taiwan and believe it is likely to continue to be relevant for China, India, parts of Africa and Latin America, among others. We believe, moreover, that there is ample room for further exploring the relationship between neo-classical insurance and classical altruism models as well as for forging better connections between the new behavioral economics and a revived application of dualism to both development theory and policy in large portions of the developing world.

NOTES

* Frank Altshuler Professor of International Economics, Yale University
The author wishes to acknowledge the substantial contribution of Michael Boozer (Yale), some early inputs from Douglas Golbin (Williams) and the assistance of Tavneet Suri (Yale), Prateek Tandon (Yale), and Zarehi Sahkyam (Yale).
\textsuperscript{1} Boeke, J. H. (1953), Economics and Economic Policy in Dual Societies, Institute of Pacific Relations.
18. IS DUALISM WORTH REVISITING?

Williamson cites such evidence of rising agricultural productivity and nearly constant real wages after the enclosure movement created a "labor surplus" condition.


Fei and Ranis, ibid.


Fei and Ranis (1997), op.cit., Appendix to Chapter 3.


Fei, John C.H., G. Ranis and Shirley W.Y. Koo (1979), Growth with Equity: The Taiwan Case, Oxford University Press.


e.g. Fields (1980), op. cit.

See Fei and Ranis (1997), op. cit.


Scott (1976) op. cit.

REFERENCES


Chapter 18

18. IS DUALISM WORTH REVISITING?

Connection.” Third Simon Kuznets Memorial Lectures, Economic Growth Center, Yale University.