Allocative Efficiency vs. "X-Efficiency"

Harvey Leibenstein

The American Economic Review, Volume 56, Issue 3 (Jun., 1966), 392-415.

ALLOCATIVE EFFICIENCY VS. "X-EFFICIENCY"

By HARVEY LEIBENSTEIN*

At the core of economics is the concept of efficiency. Microeconomic theory is concerned with allocative efficiency. Empirical evidence has been accumulating that suggests that the problem of allocative efficiency is trivial. Yet it is hard to escape the notion that efficiency in some broad sense is significant. In this paper I want to review the empirical evidence briefly and to consider some of the possible implications of the findings, especially as they relate to the theory of the firm and to the explanation of economic growth. The essence of the argument is that microeconomic theory focuses on allocative efficiency to the exclusion of other types of efficiencies that, in fact, are much more significant in many instances. Furthermore, improvement in "nonallocative efficiency" is an important aspect of the process of growth.

In Section I the empirical evidence on allocative efficiency is presented. In this section we also consider the reasons why allocation inefficiency is frequently of small magnitude. Most of the evidence on allocative inefficiency deals with either monopoly or international trade. However, monopoly and trade are not the focus of this paper. Our primary concern is with the broader issue of allocative efficiency versus an initially undefined type of efficiency that we shall refer to as "*X*-efficiency." The magnitude and nature of this type of efficiency is examined in Sections II and III. Although a major element of "*X*-efficiency" is motivation, it is not the only element, and hence the terms "motivation efficiency" or "incentive efficiency" have not been employed.

As he proceeds, the reader is especially invited to keep in mind the sharp contrast in the magnitudes involved between Tables 1 and 2.

I. *Allocative Inefficiency: Empirical Evidence*

The studies that are of interest in assessing the importance of allocative efficiency are summarized in Table 1. These are of two types. On the one side we have the studies of Harberger and Schwartzman on the "social welfare cost" of monopoly. On the other side we have a number of studies, among them those by Johnson, Scitovsky, Wemelsfelder, Janssen, and others, on the benefits of reducing or eliminating restrictions to trade. In both cases the computed benefits attributed to the reallocation of resources turn out to be exceedingly small.

* The author is professor of economics at the University of California, Berkeley.

TABLE 1—CALCULATED "WELFARE LOSS" AS PERCENTAGE OF GROSS OR NET NATIONAL PRODUCT ATTRIBUTED TO MISALLOCATION OF RESOURCES

Study	Source	Country	Cause	Loss
A. C. Harberger	*A.E.R.* 1954	U.S.A. 1929	Monopoly	.07 per cent
D. Schwartzman	*J.P.E.* 1960	U.S.A. 1954	Monopoly	.01 per cent
T. Scitovsky	(1)	Common Market 1952	Tariffs	.05 per cent
J. Wemelsfelder	*E.J.* 1960	Germany 1958	Tariffs	.18 per cent
L. H. Janssen	(2)	Italy 1960	Tariffs	max. .1 per cent
H. G. Johnson	*Manchester School* 1958	U. K. 1970	Tariffs	max. 1.0 per cent
A. Singh	(3)	Montevideo Treaty Countries	Tariffs	max. .0075 per cent

Sources:
(1) [29].
(2) [16].
(3) Unpublished calculation made by A. Singh based on data found in A. A. Faraq, *Economic Integration: A Theoretical, Empirical Study*, University of Michigan, Ph.D. Thesis, 1963.

Let us look at some of the findings. In the original Harberger study [14] the benefits for eliminating monopoly in the United States would raise income no more than 1/13 of 1 per cent. Schwartzman's [28] study which recomputes the benefits of eliminating monopoly by comparing Canadian monopolized industries as against counterpart competitive U.S. industries, and vice versa in order to determine the excess price attributable to monopoly, ends up with a similar result. Similarly, the benefits attributed to superior resource allocation as a consequence of the Common Market or a European Free Trade Area are also minute—usually much less than 1 per cent.

The calculations made by Scitovsky of the benefits to the Common Market (based on Verdoorn's data) led him to the conclusion that ". . . the most striking feature of these estimates is their smallness. The one that is really important (for reasons to appear presently), the gain from increased specialization . . . which is less than one-twentieth of one per cent of the gross social product of the countries involved. This is ridiculously small . . ." [29, p. 64]. J. Wemelsfelder [33, p. 100] has calculated that the welfare gain of reducing import duties and increasing imports and exports accordingly amounts to .18 of 1 per cent of national income. Harry Johnson in an article on England's gain in joining a Free Trade Area [17, pp. 247 ff.] calculates the net gain from trade at less than 1 per cent. That is, Johnson arrives at the conclusion that 1 per cent of the national income would be the absolute maximum gain for Britain from entering the European Free Trade Area.

A recent study by L. H. Janssen [16, p. 132] calculates that the gains from increased specialization for the different countries of the European Economic Community would be largest for Italy, but even here the amount is only 1/10 of 1 per cent of total production.[1] Janssen points out that, if the production gain for Italy due to specialization were calculated by Scitovsky's method, which he believes involves an overestimation, "the production gain in the most extreme case is still less than .4 per cent." Janssen concludes, as have others, that the welfare effects of a customs union based on the superior allocation of resources are likely to be trivial. He does, however, point to the possibility "that the mere prospect of the frontiers opening would infuse fresh energy into entrepreneurs." He recognizes that certain qualitative factors may be highly important and that the consequences of growth are certainly more significant than those of allocative welfare.

My research assistant, A. Singh, has calculated the gains from trade (following the Scitovsky method) for the Montevideo Treaty Countries[2] (Argentina, Brazil, Chile, Mexico, Paraguay, Peru, and Uruguay) and found it to be less than 1/150 of 1 per cent of their combined GNP. Even if we double or triple this result to allow for such factors as the effect of failing to take account of quantitative restrictions in the analysis, the outcome is still trivial.

Harberger's study on Chile [14] which involves the reallocation of both labor and capital yields a relatively large estimate. Harberger intends to obtain as large an estimate as possible of the consequences of reallocating resources by using what I believe to be (and what he admits to be) rather extreme assumptions in order to obtain maximum outer bounds. Despite this he comes up with a number that is between 9 and 15 per cent. However, no actual data are employed. What are used are outer-bound estimates based on personal impressions. I expect that a careful study similar to the Verdoorn-Scitovsky study would probably come up with numbers that would be no larger than 1 or 2 per cent.

The empirical evidence, while far from exhaustive, certainly suggests that the welfare gains that can be achieved by increasing *only*

[1] R. A. Mundell in a review of Janssen's book appears to reach a similar conclusion to the point made in this paper when he speculates that:

> . . . there have appeared in recent years studies purporting to demonstrate that the welfare loss due to monopoly is small, that the welfare importance of efficiency and production is exaggerated, and that gains from trade and the welfare gains from tariff reduction are almost negligible. Unless there is a thorough theoretical re-examination of the validity of the tools on which these studies are founded, and especially of the revitalized concepts of producers' and consumers' surplus, some one inevitably will draw the conclusion that economics has ceased to be important! [22, p. 622].

[2] Based on data found in [11].

allocative efficiency are usually exceedingly small, at least in capitalist economies. In all but one of the cases considered all of the gains are likely to be made up in one month's growth. They hardly seem worth worrying about.

Let us see briefly why these gains are usually small. We cannot prove that we would expect them to be small on purely theoretical grounds. If we combine our theory with what we could agree are probably reasonable estimates of some of the basic magnitudes, then it appears likely that in many cases (but certainly not all *possible* cases) the welfare loss of allocative inefficiency is of trivial significance. The idea could be developed with the aid of the diagram employed by Harberger. (See Figure 1.) In Figure 1 we assume that costs are constant

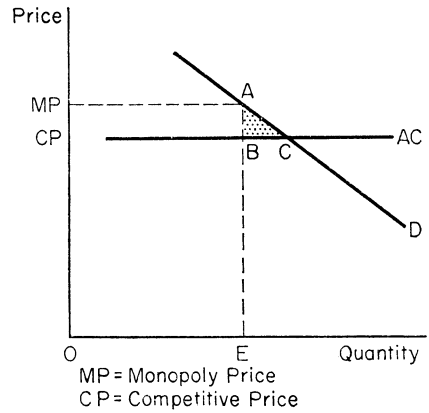

FIGURE 1

within the relevant range. D is the demand function. Under competition price and quantity are determined at the intersection C. The monopoly price is above the competitive price equal to AB in the figure. The monopoly output is determined at the point A. The welfare loss due to monopoly, which is the same as the welfare gain if we shifted to competition, is equal to the triangle ABC. We obtain an approximation to this amount by multiplying the price differential AB by the quantity differential BC by one-half and multiplying this by the proportion of national income in industries involving the misallocation.

Let us play around with some numbers and see the kind of results we get as a consequence of this formulation. Suppose that half of the national output is produced in monopolized industries and that the price differential is 20 per cent and that the average elasticity of demand is 1.5. Now the outcome will turn out to be 1½ per cent. But we really used enormous figures for the misallocation. And yet the result

is small. Monopoly prices, according to estimates, appear to be only about 8 per cent on the average above competitive prices. We can substitute some reason other than monopoly for the misallocation and still come out with similar results.[3]

Consider the cases of subsidized industries under some sort of governmental inducements to growth; and that of governmentally run industries. In the subsidy case the calculation would be similar. Suppose that as much as 50 per cent of the industries are subsidized to the extent of a 20 per cent difference in cost and that the output point on the demand function is where elasticity is unity. This last point may be reasonable since the operators of subsidized industries might want gross revenue to be as large as possible. If, on the other hand, we assume that they are profit maximizers and restrict output to a greater extent, then we might assume a price elasticity of two. This latter, however, is unlikely because monopoly profits are inconsistent with subsidized industries. Those who receive the subsidy would have the legitimate fear that the subsidy would be lowered if unusual profits were earned. Hence, behavior in the direction of revenue maximization appears reasonable and the calculated welfare loss is less than 2 per cent.

A similar result could be achieved in the case in which the government runs industries that affect 50 per cent of the national income of an economy. In all the cases we have considered, the magnitudes chosen appear to be on the large side and the outcome is on the small side.

Of course, it is possible that the magnitude of allocative inefficiency would be large if there are large discontinuities in productivity between those industries where inputs are located and those industries to which the same inputs could be moved. This, in effect. is the basic assumption that Harberger made in his study of Chile [14]. But if it turns out that there is a reasonable degree of continuity in productivity, and that the only way shifts could be made is by simultaneously increasing either social overhead capital or direct capital in order to make the shifts involved, then, of course, a great deal of the presumed gains would be eaten up by the capital costs and the net marginal gains would turn out to be rather small. My general impression is that this is likely to be the case in a great many underdeveloped countries where

[3] For the sake of completeness we should take the income effect into account in our estimation of consumer surplus. It may readily be seen that this magnitude is likely to be exceedingly small. Suppose that the initial effect of a superior allocation is 1 per cent; then the income effect for a noninferior good will be to shift the demand function to the right by 1 per cent on the average. Thus, the addition to consumers' surplus will be 1 per cent, and the consumers' surplus foregone will be roughly 1 per cent of 1 per cent. If we consider all consequent effects in a similar vein, then the estimated welfare loss will be $.010101 \ldots < .0102$. The actual magnitude will, of course, be smaller because the demand will shift to the left in the case of inferior goods. For an excellent discussion of these matters see A. P. Lerner [21].

differential productivities appear to exist between the agricultural sector and the industrial sector. One cannot go beyond stating vague impressions since there is a lack of hard statistical evidence on this matter.

Why are the welfare effects of reallocation so small? Allocational inefficiency involves only the net marginal effects. The basic assumption is that every firm *purchases and utilizes* all of its inputs "efficiently." Thus, what is left is simply the consequences of price and quantity distortions. While some specific price distortions might be large it seems unlikely that all relative price distortions are exceptionally large. This implies that most quantity distortions must also be relatively small since for a given aggregate output a significant distortion in one commodity will be counterbalanced by a large number of small distortions in the opposite direction in quantities elsewhere. While it is possible to *assume* relative price distortions and quantity distortions that would be exceedingly high, it would be difficult to believe that, without intent, the sum of such distortions should be high. However, it is not *necessarily* so on purely *a priori* grounds.

There is one important type of distortion that cannot easily be handled by existing microeconomic theory. This has to do with the allocation of managers. It is conceivable that in practice a situation would arise in which managers are exceedingly poor, that is, others are available who do not obtain management posts, and who would be very much superior. Managers determine not only their own productivity but the productivity of all cooperating units in the organization. It is therefore possible that the actual loss due to such a misallocation might be large. But the theory does not allow us to examine this matter because firms are presumed to exist as entities that make optimal input decisions, apart from the decisions of its managers. This is obviously a contradiction and therefore cannot be handled.

II. *X-Efficiency: The Empirical Evidence*

We have seen that the welfare loss due to allocational inefficiency is frequently no more than 1/10 of 1 per cent. Is it conceivable that the value of X-inefficiency would be larger than that? One way of looking at it is to return to the problem of the welfare loss due to monopoly. Suppose that one-third of the industries are in the monopolized sector. Is it possible that the lack of competitive pressure of operating in monopolized industries would lead to cost 3/10 of a per cent higher than would be the case under competition? This magnitude seems to be very small, and hence it certainly seems to be a possibility. The question essentially, is whether we can visualize managers bestirring themselves sufficiently, if the environment forced them to do so, in order to reduce

costs by more than 3/10 of 1 per cent. Some of the empirical evidence available suggests that not only is this a possibility, but that the magnitudes involved are very much larger. As we shall see, the spotty evidence on this subject does not prove the case but it does seem to be sufficiently persuasive to suggest the possibility that *X*-efficiency exists, and that it frequently is much more significant than allocational efficiency.

Professor Eric Lundberg in his studies of Swedish industries points to the case of the steel plant at Horndal that was left to operate without any new capital investment or *technological change,* and furthermore maintenance and replacement were kept at a minimum, and yet output per man hour rose by 2 per cent per annum. Professor Lundberg asserts that according to his interviews with industrialists and technicians "sub-optimal disequilibrium in regard to technology and utilization of existing capital stock is a profoundly important aspect of the situation at any time." (This according to Gorin Ohlin's summary of Lundberg's findings [24].) If a suboptimal disequilibrium exists at any time, then it would seem reasonable that under the proper motivations managers and workers could bestir themselves to produce closer to optimality, and that under other conditions they may be motivated to move farther away from optimality.

Frederick Harbison reports visiting two petroleum refineries in Egypt less than one-half mile apart. "The labor productivity of one had been nearly double that in the other for many years. But recently, under completely new management, the inefficient refinery was beginning to make quite spectacular improvements in efficiency with the same labor force" [15, p. 373]. We may inquire why the management was changed only recently whereas the difference in labor productivity existed for many years. It is quite possible that had the motivation existed in sufficient strength, this change could have taken place earlier.

In a recent book on the firm, Neil Chamberlain [5, p. 341] visualizes his firms reacting to variances between forecasted revenues and expenditures and actual. He quotes from the president of a corporation: "Actual sales revenue for the fiscal year varied one per cent from the original forecast. Expenditures varied 30 per cent. The reasons were practically entirely due to manufacturing problems of inefficiency and quality. . . . The only actions specificially taken were in attempted changes in methods of production . . . [and] the use of an engineering consulting firm. . . .' One would have thought that the cost-reducing activities mentioned could be carried out irrespective of the variance. Nevertheless, the quotation clearly implies that, in fact, they would not have been motivated to attempt the changes were it not that they were stimulated by the variance.

Before proceeding to present more empirical evidence on the possible magnitude of X-efficiency it is of importance to say something about the nature of the data. The empirical evidence does not present many unambiguous cases. Most of the evidence has to do with specific firms or, at best, industries, and not for the economy as a whole. In the evidence presented on allocative efficiency the entire economy was considered. It is quite possible that the cases considered are entirely atypical and could not be duplicated in large segments of the economy. In addition, the cases do not always deal with X-efficiency in a pure sense. Some additional inputs or reallocations are sometimes involved. Also uncertainty elements and accidental variations play a role. Nevertheless, it seems that the magnitudes involved are so large that they suggest that the conjecture that X-efficiency is frequently more significant than allocative efficiency must be taken seriously.

Now let us turn to Tables 1 and 2. In contrast to Table 1 where the misallocation effects are small, we see in Table 2 that the X-efficiency effects, at least for specific firms, are usually large. Table 2 abstracts (in the interest of conserving space) from a much more comprehensive table developed by Kilby [19] that summarizes the results of a number of ILO productivity missions. (I usually picked for each country the first three and the last items contained in Kilby's table.) It is to be observed that the cost-reducing methods used do not involve additional capital nor, as far as one can tell, any increase in depreciation or obsolescence of existing capital. The methods usually involve some simple reorganizations of the production process, e.g., plant-layout reorganization, materials handling, waste controls, work methods, and payments by results. It is of interest that the cost reductions are frequently above 25 per cent and that this result is true for a technically advanced country such as Israel as well as for the developing countries considered in other parts of the table. If the firms and/or operations considered are representative, then it would appear that the contrast in significance between X-efficiency and allocative efficiency is indeed startling. Representativeness has not been established. However, the reports of the productivity missions do not suggest that they went out of their way to work only on cases where large savings in costs could be obtained. By comparative standards (with other productivity missions) some of the results were modest, and in some cases Kilby reports that when some members of the missions returned to some of the firms they had worked on previously (e.g., in Pakistan) they found a reversion to previous methods and productivities.

There are of course a number of other studies, in addition to those by Lundberg and Harbison just mentioned which present results similar to the ILO reports. L. Rostas in his study of comparative pro-

TABLE 2—ILO PRODUCTIVITY MISSION RESULTS

Factory or Operation	Method*	Increase in Labor Productivity %	Impact on the Firm (Unit Cost Reduction)	
			Labor Savings %	Capital† Savings %
India				
Seven textile mills	n.a.	5-to-250	5–71	5–71
Engineering firms				
All operations	F, B	102	50	50
One operation	F	385	79	79
One operation	F	500	83	83
Burma				
Molding railroad brake shoes	A, F, B	100	50	50
Smithy	A	40	29	29
Chair assembly	A, B	100	50	50
Match manufacture	A, F	24	19	—
Indonesia				
Knitting	A, B	15	13	—
Radio assembly	A, F	40	29	29
Printing	A, F	30	23	—
Enamel ware	F	30	23	—
Malaya				
Furniture	A, D	10	9	9
Engineering workshop	A, D	10	9	9
Pottery	A, B	20	17	17
Thailand				
Locomotive maintenance	A, F	44	31	31
Saucepan polishing	E, D	50	33	—
Saucepan assembly	B, F	42	30	—
Cigarettes	A, B	5	5	—
Pakistan				
Textile plants	C, H, G			
Weaving		50	33	33
Weaving		10	9	9
Bleaching		59	37	37
Weaving		141	29	29
Israel				
Locomotive repair	F, B, G	30	23	23
Diamond cutting and polishing	C, B, G	45	31	—
Refrigerator assembly	F, B, G	75	43	43
Orange picking	F	91	47	—

* A=plant layout reorganized
B=machine utilization and flow
C=simple technical alterations
D=materials handling
E=waste control
F=work method
G=payment by results
H=workers training and supervision

† Limited to plant and equipment, excluding increased depreciation costs.

Source: P. Kilby [19, p. 305].

ductivity in British and American industry [26] points to the finding that differences in amount and quality of machinery per worker and the rates of utilization and replacement do not account for the entire difference in output per worker in the two countries. He further states

that ". . . in a number of industries (or firms) where the equipment is very largely identical in the U.S. and U.K., eggs, boots and shoes, tobacco, strip steel (or in firms producing both in the U.K. and U.S. . . .), there are still substantial differences in output per worker in the U.K. and the U.S." Clearly there is more to the determination of output than the obviously observable inputs. The nature of the management, the environment in which it operates, and the incentives employed are significant.

That changes in incentives will change productivity per man (and cost per unit of output) is demonstrated clearly by a wide variety of studies on the effects of introducing payments by results schemes. Davison, Florence, Gray, and Ross [7, p. 203] review the literature in this area for British industry, survey the results for a number of manufacturing operations, and present illustrative examples of their findings from a number of firms. The summary of their findings follows: "The change in output per worker was found to vary among the different operations all the way from an increase of 7.5 per cent to one of 291 per cent, about half the cases falling between 43 per cent and 76 per cent. Such increases in output, most of them large, from our 'first-line' case histories, and from additional evidence, were found not to be just a 'flash in the pan' but were sustained over the whole period of study."

Roughly similar findings were obtained for the consequences of introducing payments by results in Australia, Belgium, India, the Netherlands, and the United States [36]. In Victoria it was found that "soundly designed and properly operated incentive plans have in practice increased production rate in the reporting firms from 20 to 50 per cent." In the Netherlands labor efficiency increases of 36.5 per cent were reported. It seems clear that with the same type of equipment the working tempo varies considerably both between different workers and different departments. Appropriate incentives can obviously change such tempos considerably and reduce costs, without any changes in purchasable inputs per unit.

The now-famous Hawthorne Studies [25] suggest that the mere fact that management shows a special interest in a certain group of workers can increase output. That is, management's greater interest in the group on whom the experiments were tried, both when working conditions were improved and when they were worsened, created a positive motivation among the workers. (The magnitudes were from 13 to 30 per cent [20].) In one of the ILO missions to Pakistan an improvement in labor relations in a textile mill in Lyallpur resulted in a productivity increase of 30 per cent. Nothing else was changed except that labor turnover was reduced by one-fifth [37] [38].

Individual variations in worker proficiency are probably larger than plant differences. Frequently the variation between the best to poorest worker is as much as four to one. Certainly improved worker selection could improve productivity at the plant level. To the extent that people are not working at what they are most proficient at, productivity should rise as a consequence of superior selection methods [13, p. 147].

Although there is a large literature on the importance of psychological factors on productivity, it is usually quite difficult to assess this literature because many psychologists work on the basis of high- and low-productivity groups but do not report the actual numerical differences. In general, it seems that some of the psychological factors studied in terms of small-group theory can account for differences in productivity of from 7 to 18 per cent. The discoveries include such findings as (1) up to a point smaller working units are more productive than larger ones; (2) working units made up of friends are more productive than those made up of nonfriends; (3) units that are generally supervised are more efficient than those that are closely supervised [1]; and (4) units that are given more information about the importance of their work are more proficient than those given less information [32]. A partial reason for these observed differences is probably the likelihood that individual motivation towards work is differently affected under the different circumstances mentioned.

The shorter-hours movement in Western Europe and in the United States, especially up to World War I, has some interesting lessons for productivity differentials without capital changes. Economists frequently assume that for a given capital stock and quality of work force, output will be proportional to number of hours worked. Experiments during World War I and later showed that not only was the proportionality law untrue, but that frequently *absolute* output actually increased with reductions in hours—say from a ten-hour day to an eight-hour day.[4] It was also found that with longer hours a disproportionate amount of time was lost from increased absenteeism, industrial accidents, and so on. In many cases it would obviously have been to a firm's interest to reduce hours below that of the rest of the industry. Firms could have investigated these relations and taken advantage of the findings. For the most part, governments sponsored the necessary research on the economics of fatigue and unrest under the stimulus of the war effort, when productivity in some sectors of the economy was believed to be crucial. The actual reduction of hours that took place

[4] The empirical findings and experimental literature are reviewed in a number of places. For a brief review of the literature see [37]. See page 5 for bibliography of major works in the area.

was a consequence of the pressure of labor unions and national legislation.

In this connection it is of interest to note that Carter and Williams [4, pp. 57ff.] in their study of investment in innovations found that a high proportion (over 40 per cent) was of a "passive" character—i.e., either in response to the "direct pressure of competition" or "force of example of firms (etc.) other than immediate rivals." Unfortunately it is difficult to find data that would represent the obverse side of the coin; namely, data that would suggest the degree to which firms do not innovate for lack of a sufficient motivating force, such as a lack of competitive pressure. However, there is a great deal of evidence that the delay time between invention and innovation is often exceedingly long (sometimes more than 50 years),[5] and the lag time between the use of new methods in the "best practice" firms in an industry and other firms is also often a matter of years. Salter in his study on *Productivity and Technical Change* [27, p. 98] points to the following striking example. "In the United States copper mines, electric locomotives allow a cost saving of 67 per cent yet although first used in the mid-twenties, by 1940 less than a third of locomotives in use were electric."[6] Other similar examples are mentioned by Salter and others. A survey of industrial research undertaken by 77 companies showed that one-third were carrying on research for "aggressive purposes," but that two-thirds were "forced into research for defensive purposes."[7]

The relation between the "cost" of advice or consulting services and the return obtained has not been worked out for the ILO productivity missions as a whole. In one case (in Pakistan) the savings affected in three textile mills as a consequence of the work of the mission during the year that the mission was there "represented about 20 times the entire cost of the mission in that year." While the study does not indicate how representative this result was, the impression one gets is that rates of return of rather large magnitudes are not entirely unusual.

J. Johnston studied the return to consulting services in Great Britain.

[5] See the table in [9, pp. 305-6].

[6] [27]. See especially Appendix to Chapter 7, "Evidence Relating to the Delay in the Utilization of New Techniques." It seems to me that Salter did not quite draw the only possible conclusion from his Table 11. Plants with no significant changes in equipment, method, and plant layout had quite startling changes in output per man-hour, especially if we consider the fact demonstrated in the table that output per man-hour frequently falls under such circumstances. The range of variation in the changes (24 per cent) is larger for the plants without significant changes in equipment, etc., than for those with significant improvements. This is not to argue against the thesis that changes in techniques are important, but to suggest that significant variations in production can and do occur without such changes.

[7] See [3] for source.

For the class of jobs where it was possible to make a quantitative assessment of the results (600 jobs were involved), it was found that on the average the rate of return was about 200 per cent on consulting fees [18, p. 248]. Johnston's study is of special interest for our purposes because (a) it is a very careful study, and (b) the magnitudes of increases in productivity are of the same order (although the variations are less extreme) as those obtained in underdeveloped countries. The nature of the consulting work was not too dissimilar to that carried out by the ILO teams. On the whole they involved improvements in general management, plant layout, personnel, production procedures, selling organization, management and budgeting and accounting systems. For the consulting jobs whose consequences were quantitatively assessed, the average increase in productivity was 53 per cent, the lowest quartile showed an increase of 30 per cent, and the highest quartile 70 per cent [18, p. 273].

The studies mentioned deal with examples that are more or less of a microeconomic nature. In recent years we have had a number of studies that are their *macro*economic complements. The work of Solow, Aukrust, Denison, and others show that only a small proportion of increase in GNP is accounted for by increases in inputs of labor or capital. The "unexplained residual" covers about 50 per cent to 80 per cent of growth in advanced countries [2] [10] [23] [30] [31]. The residual comprehends a greater range of "noninput" growth factors (e.g., technological change, education of the labor force) than was covered in the examples we considered, but the motivational efficiency elements may account for some fraction of the residual. (E.g., Johnston estimates that one quarter of the annual increase in product is accounted for by consulting services.)

What conclusions can we draw from all of this? First, the data suggest that there is a great deal of possible variation in output for similar amounts of capital and labor and for similar techniques, in the broad sense, to the extent that technique is determined by similar types of equipment. However, in most of the studies the nature of the influences involved are mixed, and in some cases not all of them are clear to the analyst. In many instances there appears to have been an attempt to impart knowledge, at least of a managerial variety, which accounts for *some* of the increase in output. But should this knowledge be looked upon as an increase in inputs of production in all instances? Although the first reaction might be that such attempts involve inputs similar to inputs of capital or labor, I will want to argue that in many instances this is not the case.

It is obvious that not every change in technique implies a change in knowledge. The knowledge may have been there already, and a change

in circumstances induced the change in technique. In addition, knowledge may not be used to capacity just as capital or labor may be underutilized. More important, a good deal of our knowledge is vague. A man may have nothing more than a sense of its existence, and yet this may be the critical element. Given a sufficient inducement, he can then search out its nature in detail and get it to a stage where he can use it. People normally operate within the bounds of a great deal of intellectual slack. Unlike underutilized capital, this is an element that is very difficult to observe. As a result, occasions of genuine additions to knowledge become rather difficult to distinguish from those circumstances in which no new knowledge has been added, but in which existing knowledge is being utilized to greater capacity.

Experience in U.S. industry suggests that adversity frequently stimulates cost-reducing attempts, some of which are successful, within the bounds of existing knowledge [12]. In any event, some of the studies suggest that motivational aspects are involved entirely apart from additional knowledge. The difficulty of assessment arises because these elements are frequently so intertwined that it is difficult to separate them.

Let us now consider types of instances in which the motivational aspect appears fairly clearly to play a role. The ILO studies discuss a number of cases in which there had been a reversion to previous less efficient techniques when demonstration projects were revisited after a year or more. This seems to have occurred both in India and in Pakistan [38, p. 157]. Clearly, the new knowledge, if there were such knowledge, was given to the management by the productivity mission at the outset, and the new management methods were installed at least for the period during which the productivity mission was on hand, but there was not a sufficient motivational force for the management to maintain the new methods. The "Hawthorne Effects" are of a more clear-cut nature. Here an intentional reversion to previous methods still led to some increases in output simply because the motivational aspects were more important than the changes in the work methods. The ILO mission reports also mention with regret the fact that techniques applied in one portion of a plant, which led to fairly large increases in productivity, were not taken over by the management and applied to other aspects of the production process, although they could quite easily have done so [38, p. 157]. In a sense we may argue that the knowledge was available to the management, but that somehow it was not motivated to transfer techniques from one portion of a plant to another.

Studies which showed increases in output as a consequence of introducing payment by results clearly involve motivational elements. For

the men subjected to the new payment scheme economic motivations are involved. For the management the situation is less clear. It is possible that in many instances the firms were not aware of the possible advantages of payment by results until they obtained the new knowledge that led to the introduction of the scheme. However, it seems most likely that this scheme is so well known that this is not the case in all, or in many instances. Management quite likely had to be motivated to introduce the scheme by some factors either within the firm or within the industry. In any event, these studies clearly suggest that for some aspects of production, motivational elements are significant.

Both the ILO studies and the Johnston study speak of the need to get the acceptance of top management for the idea of obtaining and implementing consulting advice. In addition, the ILO studies make the point that low productivity is frequently caused by top management's concern with the commercial and financial affairs of the firm rather than with the running of the factory. The latter was frequently treated as a very subordinate task. Whether this last aspect involves a lack of knowledge or a lack of motivation is difficult to determine. However, it seems hard to believe that if some top-management people in some of the firms in a given industry were to become concerned with factory management and achieve desirable results thereby, some of the others would not follow suit. Johnston makes the point that, "without the willing cooperation of management the consultant is unlikely to be called in the first instance or to stay for long if he does come in" [18, p. 237]. The ILO missions make similar remarks.

It is quite clear that consulting services are not only profitable to consultants but also highly profitable to many of the firms that employ them. But it is rather surprising that more of these services are not called for. Part of the answer may be that managements of firms are not motivated to hire consultants if things appear to be going "in any reasonably satisfactory rate." There are, of course, numerous personal resistances to calling for outside advice. If the motivation is strong enough, e.g., the threat of the failure of the firm, then it is likely that such resistances would be overcome. But these are simply different aspects of the motivational elements involved.

III. *The Residual and X-Efficiency: An Interpretation*

The main burden of these findings is that X-inefficiency exists, and that improvement in X-efficiency is a significant source of increased output. In general, we may specify three elements as significant in determining what we have called X-efficiency: (1) intra-plant motiva-

tional efficiency, (2) external motivational efficiency, and (3) nonmarket input efficiency.

The simple fact is that neither individuals nor firms work as hard, nor do they search for information as effectively, as they could. The importance of motivation and its association with degree of effort and search arises because the relation between inputs and outputs is *not* a determinate one. There are four reasons why given inputs cannot be transformed into predetermined outputs: (a) contracts for labor are incomplete, (b) not all factors of production are marketed, (c) the production function is not completely specified or known, and (d) interdependence and uncertainty lead competing firms to cooperate tacitly with each other in some respects, and to imitate each other with respect to technique, to some degree.

The conventional theoretical assumption, although it is rarely stated, is that inputs have a fixed specification and yield a fixed performance. This ignores other likely possibilities. Inputs may have a fixed specification that yields a variable performance, or they may be of a variable specification and yield a variable performance. Some types of complex machinery may have fixed specifications, but their performance may be variable depending on the exact nature of their employment. The most common case is that of labor services of various kinds that have variable specifications and variable performance—although markets sometimes operate as if much of the labor of a given class has a fixed specification. Moreover, it is exceedingly rare for all elements of performance in a labor contract to be spelled out. A good deal is left to custom, authority, and whatever motivational techniques are available to management as well as to individual discretion and judgment.

Similarly, the production function is neither completely specified nor known. There is always an experimental element involved so that something may be known about the current state; say the existing relation between inputs and outputs, but not what will happen given changes in the input ratios. In addition, important inputs are frequently not marketed or, if they are traded, they are not equally accessible (or accessible on equal terms) to all potential buyers. This is especially true of management knowledge. In many areas of the world managers may not be available in well-organized markets. But even when they are available, their capacities may not be known. One of the important capacities of management may be the degree to which managers can obtain factors of production that in fact are not marketed in well-organized markets or on a universalistic basis. In underdeveloped countries the capacity to obtain finance may depend on family connections. Trustworthiness may be similarly determined. Some types of

market information may be available to some individuals but not purchasable in the market. For these and other reasons it seems clear that it is one thing to purchase or hire inputs in a given combination; it is something else to get a predetermined output out of them.

Another possible interpretation of the data presented is in connection with the "residual" in economic growth analysis. The residual manifests itself in three basic ways: (1) through cost reduction in the production of existing commodities without inventions or innovations; (2) the introduction of innovations in processes of production; and (3) the introduction of new commodities or, what is the same thing, quality improvements in consumer goods or inputs. We have ignored the introduction of new commodities, but the other two elements are pertinent here. The data suggest that cost reduction that is essentially a result of improvement in X-efficiency is likely to be an important component of the observed residual in economic growth. In addition, there is no doubt that, in some of the cases of reduced cost, new knowledge was conveyed to the firms involved, and this too is part of the residual. It is of special interest that such new knowledge involves knowledge dissemination rather than invention. The detailed studies suggest that the magnitudes are large, and hence a significant part of the residual does not depend on the types of considerations that have been prominent in the literature in recent years, such as those that are *embodied* in capital accumulation or in invention. We have considered the problem in terms of decreasing real costs per unit of output. It is clear that for a given set of resources, if real costs per unit of output are decreased, then total output will grow, and output per unit of input will also rise. Such efforts to reduce cost are part of the contribution of the residual to economic growth.

Both competition and adversity create some pressure for change. Even if knowledge is vague, if the incentive is strong enough there will be an attempt to augment information so that it becomes less vague and possibly useful. Where consulting advice is available it is significant that relatively few firms buy it. Clearly, motivations play a role in determining the degree that consulting advice is sought. The other side of the coin is that, where the motivation is weak, firm managements will permit a considerable degree of slack in their operations and will not seek cost-improving methods. Cyert and March [6, pp. 37, 38, 242] point to cases in which costs per unit are allowed to rise when profits are high. In the previous sections we have cited cases in which there was a reversion to less efficient methods after the consultants left the scene. Thus we have instances where competitive pressures from other firms or adversity lead to efforts toward cost reduc-

tion, and the absence of such pressures tends to cause costs to rise.

Some of the essential points made in the previous paragraphs can be illustrated diagramatically, if (in the interest of simplicity) we allow for abstraction from some of the realities of the situation. The main ideas to be illustrated are as follows: (1) Some firms operate under conditions of nonminimum costs, and it is possible for an industry to have a nonminimal cost equilibrium. (2) Improvements in X-efficiency are part of the process of development, and probably a significant proportion of the "residual." In what follows we assume that there are many firms, and that each firm's output is sufficiently small so as not to affect the output, costs, or prices set by other firms. For simplicity we also assume that for each firm there is an average total unit cost (ATUC) curve that has a significant horizontal segment at its trough, and that the output selected will be on that segment. When we visual-

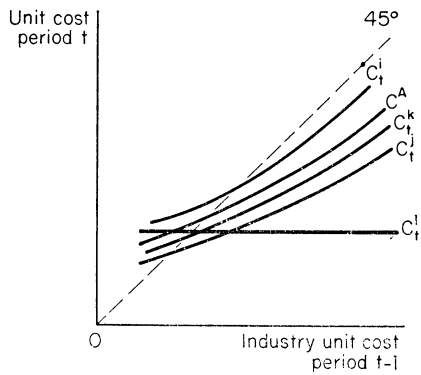

FIGURE 2

ize a firm's costs reacting to competitive conditions in the industry we imply that the entire ATUC curve moves up or down. Some firms are presumed to react to changes in the unit cost of production of the industry as a whole, i.e., to the weighted average of the unit costs of all the firms, in which each firm's weight is in proportion to its contribution to the output of the industry. Here we posit a one-period lag relation. Each firm's expectations of current industry units costs depends on actual industry unit cost in the previous period. If we choose sufficiently small periods, then this seems to be reasonable relation.

In Figure 2 each curve represents the "reaction cost line" of a firm. The ordinate shows the actual unit cost of any firm determined by that firm's reaction to what it believes or expects to be the unit cost performance of the industry as a whole. The alternate expected unit cost per-

formance of the industry is shown on the abscissa. Thus each point on line C_t^i associates the unit cost for firm i in period t, given the average unit cost in the industry in period t-1. The lines are drawn in such a way that they reflect the idea that if the unit cost that is the average for the industry is higher, then the firm's unit cost will also be higher. As average industry unit costs fall, some firms are motivated to reduce their unit costs accordingly. The higher the industry unit cost, the easier it is for any firm to search and successfully *find* means for reducing its own cost. Therefore, for a given incentive toward cost reduction, the firm is likely to find more successful ways of reducing its cost when industry costs are high compared to what they might find when they are low. As a consequence the typical reaction unit cost lines are more steeply sloped where industry unit costs are high compared to when they are low. Indeed, at very low industry unit costs the firm reaction cost lines approach an asymptote. It is not necessary for our analysis to assume that all firms are nonminimizers. Therefore some firms may have reaction cost lines that are horizontal.

The curve C^A is the average of the unit costs of all the firms in question, where the weight for any firm's cost is the proportion of its output to the total industry output. C^A is the average reaction cost line for all the firms. The basic assumption is that a firm's costs will be higher if the average industry costs are expected to be higher, and vice versa. Beyond some point, where expected average industry costs are very low, every reaction cost line will be above the 45° line.

In Figure 3 the line P is a locus of equilibrium prices. Each point on the line associates an equilibrium price with a level of industry unit cost in the previous period, which in turn determines the unit costs level of the various firms in the current period. Thus, given the industry unit cost in period t-1, this determines the unit cost level for each firm in period t. Each firm in turn will pick that output that maximizes its profits. The sum of all the outputs determines the industry output, and given the demand function for the product, the industry output determines the price. The price will be an equilibrium price if at that price no additional firms are induced to enter the industry or to withdraw from it. Thus the price for each industry unit cost is determined in accordance with conventional price theory considerations. If the price at the outset is above equilibrium price, then the entry of firms will bring that price down toward equilibrium, and if the price is below equilibrium, marginal firms will be forced to leave the industry, which in turn will cause the price to rise. Thus at every level of industry unit cost in period t-1 there is a determinate number of firms, that number consistent with the associated equilibrium price.

The point E in Figure 3, the intersection between curve C^A and the 45° line, is an equilibrium point for all the firms. The process envisioned is that each firm sets its cost in period t in accordance with its expectation of the industry cost, which by assumption is what the industry cost was in period t-1. This is a one-period lag relation. Each firm finds out what all of its competitors were doing as a group in terms of cost and reacts accordingly in the next period. If the industry cost is equal to oe then in the subsequent period each firm would set its cost so that the weighted average unit cost of all the firms would be equal to oe. Hence E is an equilibrium point.[8]

But suppose that the initial industry costs were equal to oa. We want to show that this sets up a movement that leads eventually to the point E. The firms' unit costs will average out at ab, which gener-

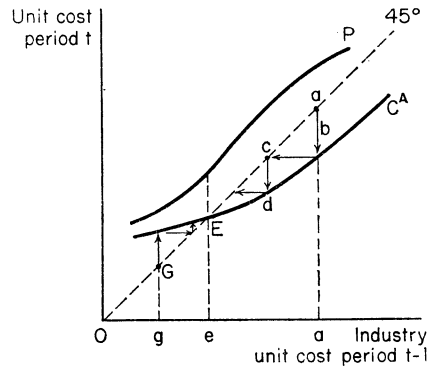

FIGURE 3

ates a process shown by the set of arrows $abcd$, etc., toward the point E. In a similar fashion, if we start with an industry cost of og, a process is set in motion so that costs move from G toward the point E. Clearly E is a stable equilibrium point. It is to be noted that every point on curve C^A need not presume that the same number of firms exist in the industry. At higher costs more firms exist, but as costs decline, some firms are forced out and fewer firms exist. In terms of the

[8] In essence the existence of an equilibrium can be shown on the basis of Brouwer's fixed-point theorem. (Point E in Figure 3 can be interpreted as a fixed point.) It would be possible to develop a much more general theory along the same lines based on less restrictive assumptions and achieve essentially the same result. For instance the one-period lag in the reaction unit cost relation can readily be eliminated. Similarly, the unique relation between the firm's unit cost and the industry unit cost level can be relaxed. See G. Debreu [8, pp. 17-18 and p. 26]. Howevêr the essence of the theory would remain the same. To conserve space and in the interest of simplicity I present the more restrictive version.

weighted average indicated by the points on the curve C^A, this simply means that some of the outputs will be zero for some of the firms as we get to lower and lower industry costs.

Figure 4 is intended to illustrate the cost reduction aspect of the residual in growth. When we begin the process the average reaction cost line is C_1^A. Firms start at point *a* and reduce costs along the arrow shown by *ab*. At this point additional information is introduced into the industry which is reflected in the diagram by the shift in the reaction cost line from C_1^A to C_2^A. Once firms are on C_2^A they then proceed with the cost reduction process as shown by the arrow *cd*. This illustrates two basic elements involved in the residual, the process of cost reduction in response to the motivation created by competitive pressures, as well as that part of cost reduction that is reflected in actual innovations, and is illustrated by downward shifts in the reaction cost lines.

IV. *Conclusions*

We have suggested three reasons for *X*-inefficiency connected with the possibility of variable performance for given units of the inputs.

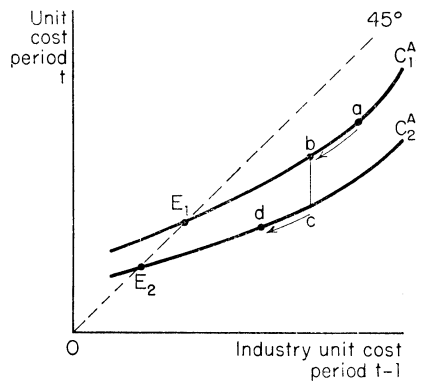

FIGURE 4

These are: (a) contracts for labor are incomplete, (b) the production function is not completely specified or known, and (c) not all inputs are marketed or, if marketed, are not available on equal terms to all buyers. These facts lead us to suggest an approach to the theory of the firm that does not depend on the assumption of cost-minimization by all firms. The level of unit cost depends in some measure on the degree of *X*-efficiency, which in turn depends on the degree of competitive pressure, as well as on other motivational factors. The re-

sponses to such pressures, whether in the nature of effort, search, or the utilization of new information, is a significant part of the residual in economic growth.

One idea that emerges from this study is that firms and economies do not operate on an outer-bound production possibility surface consistent with their resources. Rather they actually work on a production surface that is well within that outer bound. This means that for a variety of reasons people and organizations normally work neither as hard nor as effectively as they could. In situations where competitive pressure is light, many people will trade the disutility of greater effort, of search, and the control of other peoples' activities for the utility of feeling less pressure and of better interpersonal relations. But in situations where competitive pressures are high, and hence the costs of such trades are also high, they will exchange less of the disutility of effort for the utility of freedom from pressure, etc. Two general types of movements are possible. One is along a production surface towards greater allocative efficiency and the other is from a lower surface to a higher one that involves greater degrees of X-efficiency. The data suggest that in a great many instances the amount to be gained by increasing allocative efficiency is trivial while the amount to be gained by increasing X-efficiency is frequently significant.

References

1. M. Argyle, G. Gardner, and F. Cioffi, "Supervisory Methods Related to Productivity, Absenteeism, and Labor Turnover," *Human Relations,* 1958, *10,* 23-29.
2. O. Aukrust, "Investment and Economic Growth," *Prod. Meas. Rev.,* Feb. 1959, *16,* 35-53.
3. Y. Brozen, "Research, Technology, and Productivity," *Industrial Productivity*, December 1951, Industrial Relations Research Association Paper 7, p. 30.
4. C. F. Carter and B. R. Williams, *Investment in Innovations.* London 1958.
5. N. Chamberlain, *The Firm: Micro Economic Planning and Action.* New York 1962.
6. R. M. Cyert and J. G. March, *A Behavioral Theory of the Firm,* Englewood Cliffs, N.J. 1963, pp. 36-38, and Ch. 9, by O. E. Williamson, pp. 237-52.
7. J. P. Davison, P. S. Florence, and B. Gray and N. Ross, *Productivity and Economic Incentives.* London 1958.
8. G. Debreu, *Theory of Value.* New York 1959.
9. J. L. Enos, "Invention and Innovation in the Petroleum Refining Industry," in *The Rate and Direction of Inventive Activity,* Princeton, NBER, 1962, pp. 305-6.

10. S. FABRICANT, Basic Facts on Productivity. New York, NBER, 1959.
11. A. A. FARAQ, *Economic Integration: A Theoretical Empirical Study*. Unpublished doctoral dissertation, Univ. Michigan, 1963.
12. J. J. FRIEDMAN, "Top Management Faces the Cost Challenge," *Dun's Rev. and Mod. Industry*, Jan. 1961, *77*, 34-36.
13. E. E. GHISELLI AND C. W. BROWN, *Personnel and Industrial Psychology*. New York 1948.
14. A. HARBERGER, "Using the Resources at Hand More Effectively," *Am. Econ. Rev., Proc.*, May 1959, *59*, 134-47.
15. F. HARBISON, "Entrepreneurial Organization as a Factor in Economic Development," *Quart. Jour. Econ.*, Aug. 1956, *70*, 364-79.
16. L. H. JANSSEN, *Free Trade, Protection and Customs Union*, Leiden 1961, p. 132.
17. H. JOHNSON, "The Gains from Freer Trade with Europe: An Estimate," *Man. School Econ. Soc. Stud.*, Sept. 1958, *26*, 247-55.
18. J. JOHNSTON, "The Productivity of Management Consultants," *Jour. Royal. Stat. Soc.*, Ser. A., 1963, *126*, Pt. 2, 237-49.
19. P. KILBY, "Organization and Productivity in Backward Economies," *Quart. Jour. Econ.*, May 1962, *76*, 303-10.
20. H. A. LANDSBERGER, *Hawthorne Revisited*, Ithaca 1958, pp. 13 ff.
21. A. P. LERNER, "Consumer Surplus and Micro-Macro," *Jour. Pol. Econ.*, Feb. 1963, *71*, pp. 76 ff.
22. R. A. MUNDELL, review of L. H. Janssen, *Free Trade, Protection, and Customs Union*, in *Am. Econ. Rev.*, June 1962, *52*, pp. 621-22.
23. O. NIITAMO, "Development of Productivity in Finnish Industry, 1925-1952," *Prod. Meas. Rev.*, Nov. 1958, *15*, 30-41.
24. G. OHLIN, review of E. Lundberg, *Productivity and Profitability: Studies of the Role of Capital in the Swedish Economy*, in *Am. Econ. Rev.*, Sept. 1962, *52*, 827-29.
25. F. T. ROETHLISBERGER AND W. J. DICKSON, *Management and the Worker*. Cambridge 1939.
26. L. ROSTAS, *Comparative Productivity in British and American Industry*, Nat. Inst. Econ. Soc. Research Paper 13, Cambridge (England) 1964, pp. 64 ff.
27. W. E. G. SALTER, *Productivity and Technical Change*. Cambridge (England) 1960.
28. D. SCHWARTZMAN, "The Burden of Monopoly," *Jour. Pol. Econ.*, Dec. 1960, *68*, 727-29.
29. T. SCITOVSKY, *Economic Theory and Western European Integration*. Stanford 1958.
30. R. SOLOW, "Technical Progress and the Aggregate Production Function," *Rev. Econ. Stat.*, Aug. 1957, *39*, 312-20.
31. ———, "Investment and Economic Growth," *Prod. Meas. Rev.*, Nov. 1959, *16*, 62-68.
32. T. TOMEKOVIC, "Levels of Knowledge of Requirements as a Motivational Factor in the Work Situation," *Human Relations*, 1962, *15*, 197-216.

33. J. WEMELSFELDER, "The Short-Term Effect of Lowering Import Duties in Germany," *Econ. Jour.*, March 1960, *60*, 94-104.
34. S. G. WINTER, JR., *Economic Natural Selection and the Theory of the Firm.* Unpublished doctoral dissertation, mimeo, Yale Univ. 1963, pp. 83-85.
35. ANGLO-AMERICAN COUNCIL ON PRODUCTIVITY, *Productivity Measurement in British Industry.* New York 1950.
36. INTERNATIONAL LABOR ORGANIZATION, "Payment by Results," *ILO Studies and Reports,* New Ser. No. 27. Geneva 1951.
37. ———, "ILO Productivity Missions to Underdeveloped Countries, Part 1," *Internat. Lab. Rev.*, July 1957, *76*, 1-29.
38. ———, "ILO Productivity Missions to Underdeveloped Countries, Part 2," *Internat. Lab. Rev.*, Aug. 1957, *76*, 139-66.
39. ———, "Repercussions of a Reduction in Hours of Work," *Internat. Lab. Rev.*, July 1956, *74*, 23-45.