Volume Title: Six Papers on the Size Distribution of Wealth and Income

Volume Author/Editor: Lee Soltow, ed.

Volume Publisher: NBER

Volume ISBN: 0-870-14488-X

Volume URL: http://www.nber.org/books/solt69-1

Publication Date: 1969

Chapter Title: A Partial Survey of the Theory of Income Size Distribution

Chapter Author: Melvin W. Reder

Chapter URL: http://www.nber.org/chapters/c4344

Chapter pages in book: (p. 205 - 250)

A Partial Survey of the Theory of
Income Size Distribution

MELVIN W. REDER

STANFORD UNIVERSITY

I. Introduction

"THE strongly empirical tone of this chapter is of course due to the absence of a developed theory of the size distribution of income." Thus does George Stigler summarize a textbook chapter entitled "The Size Distribution of Income." [1] The validity of this remark has led me, in several places, to attempt creation where a critical recounting would otherwise have been more appropriate; where there was no theory and I thought I could provide at least a sketch of one, I have tried to do so. However, there are a good many bits and pieces of theory lying around in the literature that can, with a little trimming of edges, be fitted into a mosaic called "The Theory of Income Size Distribution"; the major part of the paper is concerned with this task.

These theoretical fragments are very disparate and sometimes are not presented in a form that indicates their relation to income size distributions. Among the arguments that do refer explicitly to income size distribution, some refer only to a single year, others to the lifetime incomes of a cohort; some refer to individuals and some to families; some to single types of incomes and some to all types combined; some to pretax and some to post-tax income, etc. If a survey of all this is to be more than a mere catalogue, items must be forced into categories with inevitable discomfort for straddlers.

The categories I have chosen are, in the logical sense, arbitrary: they have no claim to preferment beyond expository convenience. They reflect a definite view as to the role of economic theory in the present context; the purpose of an economic theory is to provide hypotheses as

to the direction and, where possible, the extent to which changes in certain structural parameters alter the size distribution of income or some component thereof. Among the more important of these parameters are those reflecting technology, tastes, and the distribution of natural abilities. Other relevant parameters are degree of resource utilization, rate of economic growth, and degree of monopoly power. It is the business of a theory of income size distribution to relate the shape of an income size distribution to these and other structural characteristics of an economy.

The emphasis of this paper is on *theory;* empirical findings are considered only to the extent that they bear upon some theory or the lack of same. But our coverage of theory is only partial. The paper concentrates on the forces that determine the distribution of the earnings of human resources, and their implications for the distribution of incomes among families. Except in passing, it does not consider the determinants of the distribution of property income among individuals and families. This omission occurs mainly because it is believed that a theory of the size distribution of property income should emphasize different forces than one pertaining to the distribution of labor income; e.g., attitudes toward risk bearing and the role of income and inheritance taxes are of much greater importance in explaining the distribution of property income. This is not to argue that a "unified" theory of size distribution is impossible, but only to suggest that near term progress is more likely if we analyze labor and property incomes separately.

Another gap in coverage arises from neglect of what has traditionally been a major concern of size distribution theorists: the characteristics of the frequency function used to describe the (size) distribution. The reason for this neglect is in no way pejorative. It is simply that much of the relevant work cannot, as yet, be related to the main body of economic theory. Where economic theory has been used to draw implications concerning the empirical characteristics of income size distributions, as in the work of Lebergott [2], Becker [3, 4] and Mincer [10, 11] the subject is given due attention.

In this paper, the income of a recipient unit is assumed to result from the sale of services—personal or property—and from government transfers. Private transfers from other recipient units, eleemosynary institutions and the like are assumed negligible. Apart from an occasional

aside, the discussion refers to societies where the income receiving unit is approximately the same as the contemporary primary family. In communities where extended kinship arrangements prevail, claims to income accruing to distant kinfolk drastically alter the meaning of asset ownership from what it is in a contemporary western society. This fact greatly complicates the comparison of distributional inequality between societies having widely different rules concerning the sharing of whatever is considered as economic goods. This is not to say that distributional questions in primitive societies are either uninteresting or unimportant, but only that economic theory in its present state has relatively little to say about them.

Present economic theory bears upon income size distribution mainly through its implications for the size distribution of rewards for productive services. If we accept the conventional, but very arbitrary distinction between an "initial" distribution of income that results from the exchange of productive services, and a "secondary" redistribution by transfers, voluntary and involuntary (e.g., taxation), then the relevance of present economic theory is confined almost entirely to explaining the initial distribution. Economic theory is relevant to analyzing the feedback effect (via incentives) of the redistribution upon the initial distribution. However, the characteristics of the redistributive mechanism itself are considered as data.

This is simply to note that economics does not yet have a usable theory to explain any of the following: the size or composition of expenditures of the public sector; the allocation of a given tax burden among income receiving units; the allocation of revenue raising as between fees and taxes. So far as income distribution is concerned, the public sector does "as the Prince wills" and the economist accepts the will of the Prince as a datum. This being the case, present economic theory is less potent the greater the role of the secondary redistribution in determining the final distribution.

Following a customary though not universal practice, I shall assume that the "fundamental" income distribution is that determined by the prices of factor services multiplied by the quantities sold. Other possible distributions are assumed to be transformations of the fundamental one, the equations of transformation reflecting the interrelation of wealth ownership (human and nonhuman) with the possession of (other)

relevant socioeconomic characteristics. This assumption is not innocuous, and to some extent it is contrary to fact.[1]

Moreover, one well-known hypothesis about income distribution, the Keynesian or widow's cruse theory, expounded by Boulding [5], Kaldor [6] and others, explicitly relates aggregate factor shares to the aggregate saving propensity which in turn is widely believed to be related to the size distribution of income. Thus, by implication, we are assuming either (1) that a family's saving propensity is not related to its income level or (2) that the relative share of family income derived from each type of factor is the same at all levels of family income, or both. (2) is patently contrary to fact (Kravis [7], p. 193) while (1) has been much debated during the last decade. However, to keep a complicated subject from becoming completely unmanageable let us assume that savings propensities and expenditure patterns generally have no effect upon the income size distribution.

Now for some brief remarks on the need to provide minimum subsistence for all members of the community. In most societies some provision is made for assuring a minimum income to every member sufficient for his survival. In time of famine or other disaster this minimum may be breached perforce, but normally it is provided. The existence of this (implicit) minimum income limits the possible income size distributions the community will tolerate.

For example, suppose a community's output of foodstuffs is F_1 and that it has n members whose average productivity in terms of foodstuffs is F_1/n and whose minimum income is F_2/n $(F_2 < F_1)$. (For simplicity assume the economy to be closed.) Then the possible size distributions must lie between complete equality at one extreme, where everyone receives F_1/n and, at the other extreme, F_2/n to the poorest $n - 1$ persons and $F_1 - [(n - 1/n)F_2]$ to the rich man. In communities where per capita food productivity is not greatly above the subsistence minimum (i.e., where F_1 and F_2 are not very different), the requirement that no one receive less than F_2/n greatly restricts the possible values of (say) the Gini income concentration ratio. For such communities, it is useful to subtract from total output the subsistence requirement, $F_1 - F_2$, and analyze the distribution of surplus, net output,

---

[1] For example, to the extent that Negroes per se are paid less than whites, their relative concentration in certain occupations causes the wage rates paid in those occupations to be lower, thereby affecting the fundamental distribution.

etc., rather than of gross output including subsistence (i.e., income).

The preceding paragraph is very Ricardian in spirit. Indeed, Ricardo's treatment of labor's share can, without substantial change, be considered as an analysis of the income share of an homogeneous class of low income receivers relative to all others, who are presumably richer. As this particular theory is quite familiar, I shall not expound it further though it may still have contemporary relevance for newly developing countries. In eschewing further consideration of the Ricardian theory, I bypass its self-conscious concern with the effect of changing the numeraire. In this matter, I follow virtually all modern writers, who consider the size distribution of income as the distribution of a magnitude measured in money.

Except for a few places where it is absolutely essential to assume otherwise, I shall argue as though the size distributions under discussion can be fully described by two parameters, mean and variance (dispersion). On this assumption, the degree of inequality in a distribution is conveniently and conventionally measured by the variance. In parts I and II, I shall discuss the income size distribution as it would be, under various alternative assumptions, in a situation of competitive long-run equilibrium. By a competitive long-run, I mean a situation in which the rewards of all factor services are equal to the values of their respective marginal physical products; the reward of any factor is the same in all of its uses; and consumption of all services and acquisition of all assets is carried to the point where marginal utilities are proportional to prices. Economic rationality of all resource owners and users is assumed. (For simplicity, corner solutions are assumed not to occur.) I also abstract from risk and uncertainty and assume that individuals behave as though they anticipated correctly and were certain that they did so. One particular implication of these conditions is that investment of all kinds, both in human and nonhuman capital, is carried to the point where its marginal product is equal to the rate of interest.

Long-run competitive conditions are always satisfied on a balanced growth path of which a stationary state is a special case arising when the growth rate is zero. Accordingly our point of departure is a competitive economy lying on some balanced growth path. Since the savings ratio has been assumed independent of the variance of the income size distribution, it does not matter what growth rate is posited.

One final assumption: in discussing long-run income distribution, I refer to the distribution of the present value of lifetime earnings. Fluctuations in income related to life-cycle earning patterns, transitory disturbances, etc., are ignored in this context. These important transitory phenomena are discussed later.

DISTRIBUTION OF TOTAL INCOME

Assume that individuals (or their families) invest their wealth so as to make its yield equal in all outlets, human and nonhuman. Given this condition, the distribution of income among recipients will be proportional to the distribution of wealth, including human capital (i.e., capitalized wage earning capacity), with the rate of return as a factor of proportionality.

Assume that at some initial moment, t_0, there is a given stock of wealth and a given distribution thereof among wealth-owning units. The wealth is valued at a given set of competitively determined long-run prices that remain unchanged throughout the time period under discussion. The ith recipient's wealth at t_0, $W_0^{(i)}$, will accumulate at a rate $W_0^{(i)} [1 + r^{(i)} - c^{(i)}]$ where $r^{(i)}$ is the rate of return to the ith recipient and $c^{(i)}$ is the rate at which he consumes his wealth. Where $r^{(i)} = c^{(i)}$, the recipient's wealth will be constant over time. His income over the interval, $t_1 - t_0$, is $W_0 \int_{t_0}^{t_1} e^{rt} \, dt$ (dropping the superscripts); his savings over the same interval are $W_0 \int_{t_0}^{t_1} e^{(r-c)t} \, dt$ and his saving ratio, the ratio of the two integrals, $1 - (c/r)$. Obviously, if some individuals have higher a's, $(a = r - c)$ than others, their wealth and incomes will grow faster and eventually the size distributions of wealth and income will approach a limit such that the ratio, $W^{(i)}/W^{(j)} \approx a^{(i)}/a^{(j)}$. I.e., if the r's are equal for all individuals, at the limit incomes will tend to proportionality with savings ratios.[2]

---

[2] At the start of this accumulation process, the approximation may be poor as it will be greatly influenced by the relative sizes of initial wealth endowments, $W_0^{(i)}/W_0^{(j)}$, but this factor will become less important with the passage of time, vanishing at the limit as $t \to \infty$. At any moment $t_g (g > 0)$, $\dfrac{W_g^{(i)}}{W_g^{(j)}} = \dfrac{W_0^{(i)} \int_{t_0}^{t_g} e^{a^i t} \, dt}{W_0^{(j)} \int_{t_0}^{t_g} e^{a^i t} \, dt}$; as t grows, $\dfrac{W_0}{\int e^{at} \, dt}$ shrinks indefinitely for both i and j, $a > 0$.

It might seem that equilibrium would require that r be the same for all individuals but this would be true only in a world in which knowledge (of all asset yields) was either free or could be acquired by all comers at the same price. Obviously, this cannot be descriptively accurate. Yet it is also true that "knowledge" of an asset's yield, once obtained, can be disseminated at a very low marginal cost, and frequently is. To be sure, wealthy individuals, wisely, spend more than poorer ones in testing and confirming information, but the poor can and do get tips on how insiders are investing. In short, the theoretical relation between the size of a recipient unit's wealth and the rate of return earned is very complicated and I shall not attempt to explore it. *Faute de mieux,* I shall abstract from differences among the $r^{(i)}$ by assuming them equal.

INHERITANCE AND FERTILITY

One factor obviously related to the long-run forces affecting the size distribution is the set of rules governing inheritance. Suppose that all heirs of a given wealth owner share equally in his estate. Also suppose that *all income,* from labor service as well as from property, derived from any asset owned by any member of the family is pooled without any effect on the incentive either to work or to save; i.e., an individual in his role as producer is the (inalienable) property of his family.

Then, abstracting from discontinuities through time in births and deaths, the growth rate of the wealth of the ith family (the ith individual and his heirs) will be $e^{at/b}$, where b is the growth rate of the family population; i.e., a/b is the growth rate of per capita wealth of the ith family. Per capita wealth and income in the ith family will be growing, stationary or declining as $a \gtreqless b$. The ratio of the wealth and income of the ith family to that of the jth family will approach $(a/b)^i/(a/b)^j$ asymptotically with the passage of time. In other words, under these assumptions, wealth and income will tend to proportionality with per capita saving ratios (remembering that all r's are assumed equal).

As soon as we abandon the assumption of equal inheritance among family members, matters become more complicated. We must distinguish between intrafamilial and interfamilial income dispersions. Suppose that each family has a set of inheritance rules that prescribes the wealth-income share of each individual in the family total, and that

each family's saving is done collectively so that it has its own saving ratio which is fixed once-for-all. Then the intrafamilial income dispersion will be determined by these rules.

By definition, income shares cannot be negative and realistically there is, for each individual, an *absolute* minimum income in any society, and a minimum income *share* in societies like ours. Thus, the lower the growth rate of the per capita savings of a family, the greater will be the fraction of the family's income preempted for meeting minimum income requirements and the smaller the part accessible to forces making for inequality.[3]

The inequality of wealth shares implied by a community's typical set of inheritance rules is obviously an important determinant of the degree of inequality in its wealth-income distribution. Given our assumptions, in the long run it would be the dominant factor if there were no correlation between a family's per capita saving rate and its institutionally determined degree of intrafamilial income inequality. However, if families with the most equalitarian inheritance rules were also those with the highest per capita saving rates, then the tendency for interfamilial inequality—measured by the variance of the distribution of per capita income among families—to increase over time, would be countered by a relative decline in the share of total income received by families with relatively great degrees of intrafamilial inequality. The net effect of these counterpoised tendencies on the per capita size distribution of both income and wealth depends upon the (asymptotic) per capita interfamilial variance of income, the set of intrafamilial income variances (among individuals) and the correlation between a family's per capita income and its intrafamilial income variance.

There are, I suspect, deep and largely unexplored institutional forces relating the dispersion of per capita income among families and the dispersion of income among individuals within a typical family. For example, where the mores that govern sharing within an extended family group are roughly equalitarian and effectively enforced, there is less need for the community to redistribute from richer to poorer individuals (which tends to reduce per capita income differences among families) in order to maintain minimum standards. This collective need is felt more strongly in periods of low (total) community income (when many

---

[3] The role of the subsistence minimum in limiting inequality in the Ricardian model (pp. 208–209) is applicable here.

persons are in danger of falling below the conventional minimum) than in more prosperous times. Consequently, while the pressure for a conventional minimum income tends to induce a negative correlation between the degrees of intrafamilial and interfamilial inequality among different communities (or within a given community over time), the strength of the correlation varies inversely with the level of community income.

A second factor might well work in the opposite direction; i.e., toward a positive correlation between the degrees of intra- and inter-familial inequality. This factor may be called the "hierarchical tendency." In communities where delegation of authority is drastically restricted and it is felt necessary to parallel "degree of responsibility" with share of income, family heads will have larger shares of family income and "important" families will have larger shares of community income than where the reverse is the case.

Thus in comparing different communities in respect of the impact upon distributional inequality among individuals of (1) inheritance rules and (2) societal rules for limiting inequality among individuals (which imply interfamilial transfers), it is important to recognize that the two types of rules are connected in a number of ways—more than specified here—but at the present state of knowledge, it is impossible to make useful generalizations. The speculations of Schumpeter [8] on the rise and fall of families are very relevant to research in this area, but they do not take us very far.[4] Clearly, further research is needed.

While this problem is intriguing analytically, it is important to remember that the argument is based on the very restrictive assumption that labor earnings as well as property income are pooled within a (potentially very much extended) family, which robs the model of direct relevance to the real world. Most likely, its main interest will lie in its use as a tool to analyze the size distribution of *property* income in isolation, with labor income analyzed separately. A start on such

---

[4] In discussing the distributional implications of inheritance rules, it is essential to bear in mind that the meaning of ownership of property may vary from one situation to another. Consider: the recipients of foundation fellowships are receiving the income of property alienated (put in trust) by its erstwhile owner. The substance of the situation might be identical, but the effect on the statistics of family income distribution very different, if the fellows were adopted by the head of the family awarding the fellowships. In other words, the meaning of income distribution varies with the nature and importance of the restraints on disposition of economic resources one "owns."

a line of analysis has been made recently by Professor Meade [9]. However, as he recognizes, his analysis abstracts from the effect of investment in human capital on the size distribution of earnings and/or on the relative shares of labor and property income. Since the size distributions of these two types of income are very different, varying their relative shares is likely to affect the size distribution of all income combined. More generally, Meade's argument simply accepts the wage share of the national income as a datum and does not attempt to explain either its size or its distribution among income size classes.

Clearly, an extensive formal analysis of the implications (for income size distribution) of the properties of various growth models is an item for the research agenda in this area. However, so far as I am aware, Meade's essay is the only work done thus far. Consequently, I shall not attempt to discuss the over-all income size distribution further, but will concentrate on the distribution of labor earnings.[5]

In Sections II and III, the discussion proceeds on the assumption that all income receiving units have only one earner and that all property belongs to him as well. (In Section IV this assumption is abandoned, in order to discuss the distribution of family income.) This particular assumption has the virtue of enabling us to treat income distribution among recipients as uniquely determined by the distribution among the heads of units, free of the important complications of multiple earners, transfer payments based on size of family, etc.

II. The Distribution of Labor Earnings in the Long Run

DEMAND CONSIDERATIONS [6]

The distribution of labor earnings, like the prices of labor services, ultimately depends upon the interaction of supply and demand forces.

---

[5] At almost the moment of going to press, I have come upon an unpublished paper by J. E. Stiglitz, "Distribution of Income and Wealth Among Individuals" (Cowles Foundation Discussion Paper No. 238, Nov. 9, 1967). A hasty perusal suggests that this contains a very elegant development of some of the ideas discussed here. Moreover, it seems clear that Stiglitz has, in a number of important respects, carried the analysis substantially further than any previous treatment (including this). So far as they deal with the same topics, this discussion and Stiglitz's seem to come to the same conclusions. However, I have not yet had sufficient time to study the paper to comment on its details.

[6] Throughout this section, and especially in the next few pages, I shall draw heavily on a previous paper [12].

However, before considering their interaction it is essential to consider the two sides of the market separately. First, let us consider demand.

To isolate the role of demand assume that the supply of labor services of every kind is exogenously determined and has zero "price" or rate of reward elasticity to any occupation or suboccupation. I.e., whatever the relative attractiveness of an occupation, there is nothing by way of training or education that any individual can do to make himself more attractive to employers within it. In other words, assume the labor supply characteristics of a population depend entirely upon exogenously determined tastes and talents. On the side of demand, assume that all employees are hired for a standard work week and that employers believe they can judge prospective worker performance with complete certainty; i.e., neither variations in weekly hours nor demand uncertainty (except where the contrary is explicitly stated) plays any role in the argument.

Given these assumptions, differences in demand prices for the services of particular individuals will determine the distribution of earnings among them. For simplicity assume that individuals differ only in respect of a single all-purpose "capacity," and are evaluated by prospective employers on the basis thereof.[7] The distribution of income will then depend upon: (1) perceived differences in output or performance among the workers; (2) the degree of consensus among the employers as to the relative merits of the performances of individuals, and (3) the marginal utility attached to superior performance by employers.

The simplest case is where performance is measured by physical output commanding a market price. In this case the income distribution among m individuals using identical quantities of complementary factors would be determined by the condition that the relative incomes of any two individuals must be proportional to their respective outputs, with product price being a factor of proportionality.[8] The distribution of earnings in this case has been discussed by A. D. Roy [13].

This case has an obvious similarity to textbook discussions of employer demand for productive services, but unfortunately it is very atypical of the real world. In most cases, the marginal *physical* product

---

[7] Alternatively this assumption might be interpreted as referring to the distribution of income or prizes in a single occupation where performance was ranked or measured by a prescribed set of rules as in a golf tournament.

[8] I am here abstracting from both monopoly and indivisibility.

of a worker cannot be measured in practice and in many cases quantification is impossible even in principle. Nonetheless, the market somehow generates pecuniary evaluations, which are inherently quantitative, and relative rewards. Directly or indirectly, it does this through the (implicit) auction of goods and services among rival claimants, even in cases where output units are heterogeneous and incomparable except in terms of selling price.

To see this, consider the difficult case of evaluating the output of painters, musicians, craftsmen, etc. While differences in output among "workers" of this type may contain a quantitative component (square inches of canvas covered or concert hours performed), there is undoubtedly a qualitative one as well, so that some artists earn much more per unit of output (measured on some nonpecuniary scale) than others. To avoid irrelevant complexity assume each painter to produce only one unit of output per time period which is consumed (physically destroyed) during the period. Also suppose that all consumers have identical preference maps, so that there is an unanimously accepted ranking of painters in order of merit (which I assume to be synonymous with desirability of possession). Further assume that prices of paintings are set by a conventional auction. Then a work—unit of output—of a higher ranked painter will always command a higher price than that of one ranked lower.[9] Furthermore, the difference between the prices paid for works of two different painters provides a measure of the difference in the value of their respective output units.[10]

Given the distribution of wealth among employers (buyers), inequality among the prices of painters' output would be increased by similarity of buyers' quality rankings. Consider first the case where there is little or no similarity in the rankings of various artists by potential buyers. In this case, "strength" of individual preference would not affect relative prices because all painters would have roughly the same composite ranking, and no individual's work would be more actively

---

[9] For simplicity, assume that all output units of a given painter sell for the same price.

[10] This measure is, of course, not the only possible one. For example, the total sales revenue from selling a specified number of output units of painter A on an "all or none" basis might exceed that from selling an equal quantity of B's output, even though (at certain quantities) the price of a unit of B's output exceeded that of a unit of A's.

sought than that of another. Clearly, if there were no similarity among buyer rankings, selling prices would be little affected by the strength of buyers' preferences among the output of competing painters. In this case, each art patron would obtain a substantial amount of consumer surplus and, assuming their output quantities to be equal, there would be little distributional inequality among the painters.

At the other extreme, where all buyers have identical tastes, the work of the most preferred painter would command the highest price, yielding him the highest income. Assume that per capita wealth and its variance *among nonpainters* is given.[11] Then, the difference between the income of the most preferred painter and that of the "second best" reflects the strength of the (universal) preference for (the amount of pecuniary difference the buying community attaches to) the work of the best over that of the second best painter; analogous statements apply to the income difference between the second and third best, etc.[12] The size of the interrank differences in painter income will increase with per capita income (variance constant) and with the variance of income among buyers (per capita level given). That is, artistic superiority is a "superior" good.

The imaginary example of the painters is of no interest in itself. However, it serves as a paradigm for a large number of situations where the distribution of income is strongly influenced by the extent of differential preference for superiority. Wherever a strong consensus on ranking exists and a value is attached to superior rank, rewards will tend to parallel rank. Cases where this relationship is very pronounced arise among professional athletes where the rules of the game map a set of performances into a rank order; sometimes (e.g., in professional tennis and golf) a specific monetary payoff is attached to each rank. Game type situations (where winning, losing and ranking are reasonably objective) tend to generate substantial degrees of inequality among the participants. These situations include law (barristers) and commission selling of all kinds, as well as participation in games per se.

---

[11] The reason for excluding painters is to enable us to treat the effect of wealth changes of art buyers as exogenous, which simplifies the discussion without obscuring matters of substance.

[12] This sentence expresses a definition; not an empirical statement. The strength of an individual's preference for A over B is *defined* as how much more he would pay for the (only) unit of A than for the unit of B.

In game type situations, winning may be valued for itself, but more often it serves as a means to pecuniary ends. The champion usually draws a larger paying audience than other players; the clients of a more successful lawyer, or the employer of a better salesman, have a greater expected gain (gross of compensation to the lawyer or salesman) than those of one less successful. Where differential performance is directly related to pecuniary returns, the distribution of rewards is similarly related. But where differential achievement does not have a specific monetary effect, more complicated relations may arise; e.g., rewards may be related to rank order of performance; to "quality" of performance (regardless of rank) or to both.

In nongame situations relative performance may be difficult to rank, e.g., in teaching. Difficulty in ranking is, of course, related to lack of generally accepted standards of performance, though the two are really different matters: i.e., there might be generally accepted methods of ranking, but disagreement as to desiderata. Where rankings differ widely among prospective employers, demand tends to be spread more or less evenly among candidates for (say) teaching posts, tending to equalize teacher earnings.

Because the rank order of teachers' performances cannot be determined operationally, institutions cannot be sure of the relative quality of their employees and by the same token cannot be humiliated by being caught in the act of employing *somewhat* "inferior" ones. Consequently, those employers of teachers who are concerned with the reputation of their pedagogues, rather than with their actual performance, will substitute cheaper ones whenever the adulteration of the product cannot be detected. Other employers, pressed hard by budget restraints, will find vague or contradictory rankings an excuse to opt for lower-priced teachers: "one isn't sure that they are really inferior and, even if the worst is true, it is not at all clear how serious are the consequences."

Ignorance of relative merit is not the same as indifference to it, but one fosters the other and their combined effect acts as a serious brake upon attempts to attract superior teachers by competitive bidding.[13] The

---

[13] The relation between knowing and caring about rankings is complicated. However, I doubt that it will be disputed that knowing a rank order—given some minimal concern—increases the effort to rise within it. George J. Stigler, *Trends*

above remarks refer to "pure" teachers. Where research is involved, publication facilitates comparison, and the grant obtaining process assigns dollar magnitudes to (some aspects of) differential success.

The reason for these rather discursive remarks on the relation of ranking to the determination of demand prices for qualitatively different services is the lack of a suitable general purpose scale for measuring value differences in heterogeneous economic goods. The theory of demand is, after all, a matter of the interrelation of prices and quantities of homogeneous goods. Therefore, it is natural for economists concerned with this problem to seek parameters which reflect differences in demand intensity in situations where differences in marginal products are not conducive to quantification.

One parameter which a number of economists have seemed to find useful may be termed, following Mayer [15], the "scale of operations" effect. The cluster of ideas underlying this concept has been discussed by Simon [14], Jacques [16], Lydall [17, 18], Tuck [19] and Reder [12], as well as by Mayer.

To fix ideas, think of every human agent, self-employed or wage earner, as manager of a block of resources, human or otherwise. Assume that within each job class or occupation there is unanimity as to the rank ordering of individuals with reference to their desirability as employees; also assume that nonpecuniary preferences exercise no effect on choice of employment. Then to get a superior employee it will be necessary to pay more than to obtain one who is inferior. Consequently, in equilibrium, the rank of employees in terms of desirability will correspond to their rank in the earnings hierarchy.

The hypothesis is that the most desirable employees will also be those to whom it is found profitable to assign the largest "scale of operations." I would suggest that an individual's scale of operation be measured by the (total) value product of the resources he manages; i.e., the value of superior performance (to an employer) on a given job increases with the value of the resources the job holder manages. Most writers have discussed an individual's scale of operations in terms of the number of his organizational subordinates rather than the value of the resources

in *Employment in the Service Industries,* Princeton for NBER, 1956, pp. 130–132, argues in a somewhat similar vein concerning the earnings of teachers vis-a-vis independent professionals.

he manages. However, there is no reason to ignore nonhuman resources in this context; diamond cutters are paid more (per hour) than cutters of less precious materials; operators of more expensive machinery more than those of less expensive. Indeed, as Mayer [15] and Jacques [16] point out, the relation of compensation to degree of "responsibility" is a widely accepted principle of job evaluation.

To ward off any possible squabble over claims to priority in discovering this indicator of demand intensity, permit me to quote from a little-known eighteenth century precursor:

> Fourthly, the wages of labour vary accordingly to the small or great trust which must be reposed in the workmen.
> The wages of goldsmiths and jewellers are everywhere superior to those of many other workmen, not only of equal, but of much superior ingenuity, on account of the precious materials with which they are intrusted.
> We trust our health to the physician; our fortune and sometimes our life and reputation to the lawyer and attorney. Such confidence could not safely be reposed in people of a very mean or low condition. Their reward must be such, therefore, as may give them that rank in the society which so important a trust requires. The long time and the great expense which must be laid out in their education, when combined with this circumstance, necessarily enhance still further the price of their labour.[14]

The empirical evidence to support the existence of a scale of operations effect is, at best, scattered. One of the more important pieces (of evidence) concerns the relationship among compensation of corporate executives, current sales of their companies and current profits of their companies. It appears that the partial regression of compensation on sales, profits constant, is much stronger than that of compensation on profits, sales constant, though both partial regression coefficients are positive. This relationship has been found by Roberts [22], Patton [23] and McGuire, Chiu and Elbing [24].

This particular finding implies that the differential contribution of a superior executive to a firm (over that of an inferior one) is more closely related to a firm's scale of operations (measured by sales) than to its current profits. That is, superior management may be manifested by minimizing (unavoidable) losses as well as by earning current profits. A positive association of executive compensation with sales (scale of

---

[14] A. Smith, *The Wealth of Nations,* Book I, Chapter X, pp. 93–94, Everyman's Library, 1910.

operations) is what would be expected if differential compensation were related to differential skill via the scale of operations effect.

In principle, this association should be found among employees of any rank. Unfortunately, detailed information is not readily available on compensation of employees within given occupations by the volume of resources managed. It is generally believed that wages vary among firms, with capital per head, and the results of empirical studies of (aggregate) production functions confirm this relation in the large. But available evidence is not, as yet, very impressive so far as the present argument is concerned.

Nevertheless, I am inclined to believe that the scale of operations is an important determinant of relative demand prices for various types of employees. If so, one can say that within an occupation or job class, the dispersion of demand prices for workers varies with the dispersion of gross value product per worker. However, scale of operations is not the only parameter affecting these relative demand prices; there is also "sensitivity."

The sensitivity of a job varies with the concern an employer has with the performance of its holder. Formally, we might define the sensitivity of a job as some indicator of the range of its possible value products, the actual value product attained varying with the skill of the job holder, the value of complementary resources constant; unfortunately, this does not help much. It is perhaps more suggestive to designate a job as routine or insensitive (having low sensitivity) if there is but a negligible marginal product to a higher degree of any aptitude beyond that commonly possessed by persons filling such jobs. Bank teller is a good example of a job having low sensitivity though utilizing a large volume of complementary resources. Despite numerous jokes about the dangers of abscondence, banks do not seek unusually honest men or offer them high salaries to induce them to become tellers. Character references, a bond and minimum standards of aptitude and education suffice.

Examples of jobs with high sensitivity are those of trial lawyer (barrister) or surgeon. Superior trial lawyers get cases involving larger sums than their inferiors, and top-notch surgeons operate upon wealthier bodies than the average. This might suggest that we are still observing the scale of operations effect. But this is only partially true: solicitors handle cases of roughly the same range of values as barristers,

yet their earnings dispersion is much lower.[15] Similarly, internists, psychiatrists and other medical specialists treat both rich and poor, but their income dispersion is generally believed to be less than that of surgeons. These differences in dispersion I attribute in large part to different degrees of concern for superiority and hence to greater concentration of demand on the best people (with consequent greater dispersion of fees and incomes) among surgeons and barristers than among others. This differential degree of concern for superiority is what is meant by different degrees of sensitivity among jobs. Adam Smith (in the passage quoted above) lumped scale of operations and sensitivity into "trust." [16]

Other cases where sensitivity visibly affects the dispersion of earnings occur where star performers (athletes, concert musicians, actors, etc.) earn more than their managers but other performers earn less. The conflict between sensitivity and responsibility (scale of operations) is also observed in civil service and other formal organizations, especially in universities, where the market for star researchers sometimes requires

[15] See Royal Commission on Doctors' and Dentists' Remuneration: 1957–1960; Command Paper 939, Her Majesty's Stationery Office, London, 1960, Table 7, p. 37. The upper quartile is 193 per cent of the median for barristers; 155 per cent for solicitors; the highest decile is 358 per cent of the median for barristers, but only 235 per cent for solicitors.

[16] Though the Smithian argument is appealing to common sense and hardly new, it has never been incorporated into the main part of price theory. This is no place to work out the details of an appropriate theoretical development, but its general lines would be as follows: Assume that higher quality workmen have a higher supply price to an employer. The reason for paying the higher price necessary to obtain superior workers to process a more valuable material is the same as that for taking greater pains (spending more) to guard objects of greater value.

Assume there is a known probability of dollar loss—through damage—to an object, which is an increasing function of the object's value, and a decreasing function of the amount spent to protect it. (I abstract from loss through theft in order to avoid complications of a game theoretic nature.) Assuming the owner (or insurer) of the object maximizes expected utility, he will spend an amount on protection that makes the reduction of expected dollar loss (marginal gain) from added protection equal to the marginal cost of providing protection. The reduction in expected dollar loss from a given increase in protection will increase with the value of the object but the marginal cost of any given amount of protection is independent of the object's value (or varies only slightly with its value) so that at a level of protection at which marginal gain equals marginal cost of protection for a less valuable object, it exceeds it for one of greater value and hence protection is pushed further. To hire better quality workmen is to provide more protection for materials; assuming that higher quality workmen have a higher supply price, this implies greater marginal cost.

that they receive higher salaries than their own organizational superiors as well as of others "responsible" for resources of much greater value. That the conflict between market and organizational imperatives is not always resolved in favor of the former does not gainsay the point under discussion.

In fine, both scale of operations and degree of sensitivity are parameters related to differential demand prices for workers' services. The former has the decided advantage of frequently being measurable independent of its effect on income while the latter, in general, is not. However the relevance of the latter factor should not be overlooked.

SUPPLY CONSIDERATIONS

The discussion of the previous section was concerned only with causes and effects of differences in demand prices on differential earnings in various occupations. That it may be suggestive of certain observable phenomena indicates that we are dealing with situations where, for one reason or another, lower earners are unable to change their characteristics so as to become very close substitutes for those earning more. However, if sufficient time is allowed, there is a good deal of supply elasticity to most sets of personal characteristics that yield unusually high earnings, and we must take this into account.

In general, if a particular degree of a particular capacity is highly rewarded, investment in training can often duplicate—or at least approximate—it, thereby increasing its relative supply and reducing its relative reward. Indeed, earnings associated with a given degree, y, of any capacity, A (call it A_y) cannot exceed those of B_z (or A_z) by more than the marginal cost of transforming (training) persons with B_z to A_y.

To appreciate the role of training in limiting the dispersion of earnings, suppose a given set of employer demands confronts a given set of native talents.[17] Then, in long-run equilibrium the lower the elasticity of substitution among different kinds and qualities of service, and the greater the dispersion (among individuals) of degrees of each kind of

---

[17] Native talent is obviously not an easy concept to define, and its proper definition varies with the problem. For the present purpose, we may suppose it to be the capacity of individuals at any age at which further training or education occurs outside the family. Obviously, this definition ignores all-important differences in training within the family whose effects are confounded with those of genetic factors.

talent, *ceteris paribus,* the greater will be the dispersion of earnings.[18] In a given disequilibrium situation, assume that the distribution of earnings is such that some earning difference, e.g., that associated with $A_y - B_z$, exceeds the marginal cost of transforming (through training) a lower earner into a higher. In order to reach equilibrium all such opportunities must be exploited lowering the earnings of the A_y's and raising that of the B_z's, thereby making the dispersion of earnings less than if such training were technically impossible. The greater the price of resources used in training and the greater the physical quantities needed to accomplish given results, the fewer will be the "scarce talents" subjected to competition from trained substitutes, and the fewer the individuals whom it will pay to train in order that they may compete with higher earners. Hence the greater the general level of training costs, the dispersion of native talents given, the greater the dispersion of earnings.

In practice, differences in earnings reflect both differences in the amount of capital invested in training and in native talent. If native talent were the same for all individuals; if they all had the same tastes and could obtain capital for training purposes on the same terms, then in long-run equilibrium earnings would be the same for all.

It is probable that relatively high costs of securing capital for training are associated with relative deficiencies in native talent,[19] and also that perceptions of relative deficiencies in learning capacity tend to discourage investment in training the less talented. For both reasons, it is likely that there will appear to be a positive association of amount of education with both native talent and earnings.

Differential capacity of individuals (or their families) to finance or otherwise obtain access to training has always been an integral part of

[18] If employers generally accept indefinitely inferior services as substitutes for superior ones because of arbitrarily small differences in price, the dispersion of earnings will be zero (in long-run equilibrium), no matter how great the dispersion of talent. Conversely, if all individuals have the same native talent and training, the dispersion of earnings will be zero no matter how anxious employers are to obtain superior quality.

[19] This may be due to a correlation of genetically associated incapacity of offspring with low wealth of parents. Relatively poor parents, because of their poor credit standing, can borrow for training of their offspring only at relatively high rates of interest, if at all, and hence are deterred from doing so. But irrespective of genetic factors, capacity to benefit from training is strongly associated with home environment so that children of low income families—who "can't afford to invest in training"—appear from an early age to be relatively deficient in intellectual capacity.

the intercorrelation of training, talent and earnings. Consequently, it is reasonable to suppose that a reduction in the difference in marginal cost of borrowing for training purposes (as between any pair of individuals) would alter their relative earning capacities. That is, a reduction in the dispersion of borrowing costs among individuals would tend to increase the relative amounts of training undertaken by the less talented (who are also the low earners), which would tend to raise their earnings relative to those having an initial advantage in acquiring training.

However a further analytical consideration, and one with considerable practical importance, complicates the relation between unequal access to training and unequal earnings. The marginal return imputable to native talent relative to that imputable to training, will surely vary as investment in training becomes greater and more widespread. As appropriately trained persons become more abundant, jobs which have paid well heretofore because of the great training required, but which require little native talent, will decline in earning power relative to those requiring greater degrees of native talent.[20] If it should happen that those individuals who initially received more training, because of easier access, should also have more native talent than others, then the effect of equalizing access to training might increase the dispersion of earnings and not the reverse. That is, as the relative advantage (of the initial high earners) from greater training diminishes, their earning advantage from greater native ability may increase sufficiently (because of its increased scarcity relative to investment in training) to overbalance the effect of a more equal distribution of training.[21]

The practical importance of this point stems from the worldwide trend toward more equal opportunity for education and training, at least within industrialized countries. On grounds both of ethics and efficiency, I approve of this trend. However, it is by no means clear that its consequences will be egalitarian, that it will make the distribution of

[20] It is important to note that this argument implies both an increase in total investment in training and more nearly equal access to what training there is. It is a combination of both that is usually intended by the phrase "more equal opportunity for education," but it is nonetheless important to distinguish between the two. It is unlikely that merely transferring "education" without increasing the total "amount" available would have the effects usually expected of broader educational opportunity.

[21] Though unlikely, it is possible that this could occur even though the initial correlation between training and native ability was negative.

earnings more equal. As indicated in the previous paragraph, it is possible—though by no means certain—that reducing differences in access to opportunities for training (combined with an increase in the general level of training) will *increase* the dispersion of earnings. A "meritocracy" may be more just than an aristocracy, but it is not clear that it will be more egalitarian.[22]

In the discussion thus far we have treated native talent as though it had but one dimension. This obscures certain problems as well as being patently false. Accordingly, let us consider a two-dimensional (two-talent) concept; i.e., let native talent be a vector of two components. The difference between the one- and two-talent cases arises from the possibility of various correlations between the talents. If the correlation between the two is very strong and positive, the two-talent case degenerates to the one-talent. But if the correlation is strongly negative and the two talents are of roughly equal economic importance, this case will exhibit less dispersion of earnings than that of one talent. For example, suppose that earnings are an increasing function of each talent and the talents are scaled so that earnings are a function of the sum of the "scores" on each. Then, if the summed scores of all individuals were equal (implying, inter alia, negative correlation between the rank of an individual in respect of one talent and his rank in respect of the other) earnings would be distributed equally. Although negative correlation between the two talents is an equalizing factor, it is not hard to construct examples where moving from a one-talent to a two-talent situation would increase the dispersion of earnings even though there were negative rank correlations of these two talents among individuals.[23] Increasing the number of talents beyond two does not appear to add any substantively interesting results.[24]

----

[22] This point is analogous to an argument of A. D. Roy [25]. It is also made by Meade [9], p. 61. I am indebted to E. C. Budd for the Meade reference.

[23] For instance, consider the case where the rewards are paid in the form of $1 for each "point" on each capacity, and both capacities are distributed normally with negative correlation between them, but with different variances. Suppose that initially, earnings depended only upon the capacity with the smaller variance. Then whether introducing the second capacity into the payoff mechanism decreases dispersion will depend upon the relative size of the two variances and the strength of the (negative) correlation. If the variance of the second capacity is enough larger than that of the first, earnings dispersion may increase even with perfect negative correlation.

[24] For further discussion of the role of different capacities in determining relative earnings see Tinbergen [20] and Mandelbrot [21].

Another important aspect of the earnings distribution is revealed by examining the case where "talent" is identified with ability to work in a particular occupation. Assume, for simplicity, that an individual can practice only one occupation at a time, and that the dispersion of talent for occupation A is greater than for occupation B where talent is measured in "points." Then an increase in demand for the services of A relative to B will increase the value of a point of A talent relative to a point of B talent thereby increasing the over-all dispersion of earnings; the reverse will apply where there is an increased relative demand for B services. Clearly the over-all size distribution of earnings will be affected by the pattern of demand for labor. The greater the tendency for employers to demand those services the talent for whose production (allowing for training) varies greatly among persons, the greater the dispersion of earnings, and vice versa.[25]

GENERAL EQUILIBRIUM

The discussion in the demand section ran in terms of variables related to capacity to produce desired results, but in the supply section the argument was concerned with various types of "talent" presumably measured by performance on tests. A theory of income size distribution consistent with (general equilibrium) theory requires that a set of transformations be established between the units of "talent" that measure supply and the units of "capacity" that measure demand. Needless to say, I cannot specify such a set of transformations nor so far as I am aware can anyone else at this time. This deficiency is serious.

Economic theory evades this problem by classifying productive services into factors of production each of which has an equilibrium price. But this expedient makes the question of determinacy a crucial one. Quite apart from matters of functional form, it is a fundamental question as to whether there is a set of classes of productive services (factors) whose number and dimensions are such that both the units of supply and the units of demand can be transformed into them, so as to yield determinate and consistent price-quantity relations.

One possible approach to this problem is suggested by Tinbergen [20]. He proposes (in effect) to replace a set of discrete factors of

[25] I do not elaborate on this point, despite its importance. One reason is that it would take us too far afield; also it is well discussed, though very briefly, by A. D. Roy [25].

production by a continuous function of many attributes of personal capacity; different factors of production would then be represented by different positions on the function (i.e., by different combinations of attributes). Whether this procedure will solve the problem depends on whether it is possible to find a suitable set of attributes [26] and to define supply and demand functions upon them such that consistent and determinate price-quantity relations will result.

Personal attributes such as intelligence and physical strength are better suited for measuring supply than demand. What employers seek is "productive capacity" which is separated from "personal attributes" by a thick curtain of sociopsychological variables conveniently labeled incentive, attitude and experience.[27] It is safe to say the role of these variables in the theory of wages is not well understood. Scale of operations and sensitivity are interesting correlates of labor demand; they are important indicators of the demand prices for (successful) applicants for particular *jobs* but they are not useful indicators for analyzing the demand for the services of *individuals*.

The distinction between the demand price for jobs and that for the services of individuals is important because job characteristics are determined simultaneously with wage rates: job characteristics should be endogenous in a general equilibrium model. For example, if managers do

[26] It may be that there is none. For example, employers may, at *initial* hiring levels, demand personal attributes definable in terms of objective criteria. However, subsequent advancement may depend also upon past performance as perceived and valued by the employer or his representatives. This would imply that employer demand functions were defined in terms of both personal attributes *and* past performance. A useful theory of labor demand would then have to include an employer response mechanism to past performance. The "Theory of Rank," a la Tuck, Simon, Lydall, *et al.*, might be viewed as a first—though excessively "static"—approach to such a theory.

[27] Tinbergen's discussion relates primarily to the supply side. If employer demand were for the same attributes—or some transform thereof—as workers supply, the problem would be solved. But, in general, employers are not interested in the physical strength, dexterity, intelligence, etc. of their employees except insofar as these are indicative of performance. And it is at least possible that the characteristics employers seek—consciously or otherwise—cannot be consistently transformed into those that workers supply. If this is the case, imbalances between supply and demand (either in units relevant to supply or in those relevant to demand) have to be adjusted by nonprice rationing and/or by adjustments of characteristics demanded or offered, as well as of wage rates. If so, institutional rules governing these adjustment procedures would have to be built into the theory as an integral part. I shall discuss this possibility in a short-run context in Section III; its long-run relevance is moot.

not vary greatly in capacity, neither will firm sizes nor their salaries. That is, equality of managerial skill will encourage an industrial structure of numerous small firms, each firm manager having relatively few subordinates. Conversely, large differences in managerial skill will promote an industrial structure conducive to large firms (each employing many workers) and to a correspondingly large variance of salaries and scales of operation among the members of the labor force.

Of course, the distribution of managerial capacities does not operate alone; it interacts with a set of technological possibilities, transportation costs, etc. to determine firm size, income distribution, etc. But the direction of its thrust is clear. Similarly, if reliable workers able to utilize delicate machinery are scarce, and the damage done by inadequate workers great then, technology given, the economy will tend to specialize in products, production methods and organizational forms that minimize capital per worker thereby limiting his scale of operations. By making the number of an executive's subordinates a "technological" datum instead of an economic variable to be determined by the system, Lydall, Tuck, *et al.,* improperly exclude supply factors from the picture.

HUMAN CAPITAL

The above discussion portrays the theory of income size distribution as a cluster of imperfectly joined, ill-fitting pieces. Superficially, at least, this view is quite different from the bold and imaginative picture presented by Gary Becker [4]. Becker's model is, in principle, one that explains the distribution of earnings as resulting from the investment of differing amounts of capital in the various members of the labor force.

In essence Becker assumes each individual to invest in himself a quantity (measured in dollars) of training and education such that the marginal rate of return on investment equals the marginal rate that must be paid to obtain funds. Each individual has a demand curve (D) for investment in his own capacity which relates the amount invested to marginal rate of. return, and a supply curve (S) of funds (for such investment) that relates marginal cost of borrowing to amounts borrowed. Becker shows quite easily that to obtain a finite amount of investment, either D must slope downward or S upward, or both.

If two or more individuals had identical D and S curves, they would make the same investment and have the same earnings; if all individuals

had the same demand curve (i.e., identical abilities) but different supply curves, then investment and earnings would be greater the lower the individual's borrowing costs. Conversely, if all individuals had the same supply curve but different demand curves, investment and earnings would be greater for individuals with greater abilities. In the general case, both supply and demand curves differ among individuals, and the resulting distributions of investment and earnings depend upon intercorrelations of the parameters of these curves (i.e., upon whether those with greater ability also face lower borrowing costs or the reverse) as well as upon their elasticities.

While this analysis, to which I cannot do justice in a few paragraphs, is valid within a properly restricted domain, it does not provide a basis for a generally applicable theory of the size distribution of earnings. The principal difficulty is that the shapes of the demand and supply functions are not given independently of the set of equilibrium prices, but reflect and vary with them. This can lead to situations where as between two positions of static equilibrium, A and B, A has a greater variance of human capital than B, but B has a greater variance of earnings. Thus moving from A to B would increase the variance of earnings but decrease that of human capital.

Consider the following possibility. The S curves of most individuals become more elastic, in the vicinity of their previous equilibrium positions, while no one's S curve becomes less elastic. Given all D curves, it is plausible to suppose that this rotation of S curves causes the distribution of human capital to become more equal, though this result is not logically necessary. But it is quite possible that this would lead to an absolute reduction in the supply of untrained workers, reducing outputs and raising prices of commodities in whose production they specialize. Suppose that the dispersion of native talent for unskilled work is much greater than for other types, and that differences in capacity for unskilled work cannot be greatly reduced by training. Then, if the price elasticity of demand for the output of the "unskilled sector" is sufficiently low, and the dispersion of native talent in producing the output of this sector sufficiently great, the effect of drawing off part of its labor supply will be to increase over-all dispersion of earnings despite the more equal distribution of human capital.

In effect, what I am saying is that altering the shapes of the supply curves of funds (for investment in human capital) for a large number of individuals may so alter relative prices of products and native talent as both to *lower* some individuals' demand curves for human capital (by raising the earnings those with great native talent must forego in order to receive training) and simultaneously to increase their earnings.[28] For some of these individuals, the *D* curves might even have negative ordinates over some part of the investment axis. This result could arise not only from the aforementioned shift of *S* curves, but also from a shift in tastes away from the output of trained workers toward that of untrained workers, at given prices and incomes.

This is not to argue that the distribution of human capital is unrelated to the distribution of earnings. Rather it is to say that both distributions are determined simultaneously by an underlying set of consumer tastes, productive techniques and wealth ownership. The distribution of wealth greatly influences the shapes and positions of the *S* curves. These underlying variables determine the positions and interrelations of the *D* and *S* curves through relative prices. They also determine the size and distribution of the stock of capital (human and otherwise) and the distribution of earnings. Any particular relationship between the size distributions of human capital and of earnings is conditional upon the interrelation of these underlying exogenous variables.[29]

One further point made by Becker especially merits comment. He argues that in the absence of fairly strong negative correlation between ability and marginal borrowing costs, ability and self-investment will be positively correlated, even if both were symmetrically distributed. The purpose of this is to show how a symmetrical distribution of ability can be reconciled with a skewed distribution of earnings. It has often been argued that any satisfactory theory of income size distribution must make this reconciliation because abilities are distributed symmetrically while·

[28] At the outset of his paper, Becker assumes that no part of a worker's income is return to native ability, which would preclude this possibility. However, this assumption is arbitrary, as he admits.

[29] It should be obvious that the spirit of these remarks is quite similar to that in which Mrs. Robinson has quarreled with users of the concept, "quantity of capital." However my intent is less destructive than hers sometimes appears to be; I insist only that users of *S* and *D* curves consider whether these curves remain unaltered (and if not how they are altered) by changes in exogenous variables or functional forms that cause changes in relative equilibrium prices.

the size distribution of earnings is positively skewed. I offer the following remarks on this contention:

(1) Not all income size distributions are positively skewed, or skewed at all. Staehle [26] showed many years ago that, among manual workers, distributional symmetry of earnings is approached quite closely once the effect of unemployment is eliminated. Miller [29, pp. 18–28 and 58–63] also presents data from the 1950 Census to the same general effect. This is not to deny that the over-all earnings distribution is skewed, but to raise a question as to the range of applicability of a theory that implies skewness.[30]

(2) It is only a *complete* theory of income size distribution that must reconcile a symmetrical distribution of abilities with a skewed distribution of earnings. Becker's theory does not pretend to completeness as it abstracts from risk and uncertainty. Friedman [27] has argued that skewness can result solely from phenomena related to differential attitudes concerning risk bearing. If so, there is no need for Becker's argument; indeed, it is necessary to show that the simultaneous operation of the forces operating in Becker's model, together with those operating in Friedman's, generates a plausible degree of skewness—neither too much nor too little.

(3) Lydall [18], Simon [14] and others have also provided rationalizations of skewed earnings distributions by means of the scale of operations effect, as well as other explanatory variables. Mincer [10] provides two distinct rationalizations of the skewness of aggregate earnings distributions. To the extent that their arguments are valid they must be integrated with one another, and with Becker's, to show that a *complete* model generates an empirically relevant functional form for the earnings size distribution. So long as a theory does not purport to be complete it is not necessary that it make the aforementioned reconciliation; it is sufficient that it not preclude such a reconciliation being made by other parts of the model.

---

[30] On this point, an earlier version of Becker's theory [3, p. 65] very neatly explains Miller's findings [29] that skewness is found mainly in occupations requiring a great deal of education. The argument is that earnings are a weighted sum of return to (symmetrically distributed) unskilled ability and to investment which is asymmetrically distributed; the investment component is greater in occupations requiring more schooling. This feature of Becker's theory has, for some reason, been abandoned in the later version [4].

III. Life Earning Patterns and Social Stratification

In the previous section, we have ignored the well-known association of age with earnings. However, theorists of income distribution have been very anxious to explain life-earning cycles and especially differences in these cycles for persons at different levels of education. We now turn to this subject. The argument of this section refers mainly to short-period variations in income distribution, e.g., year to year, and not to those longer periods in which variations in the distributions of wealth and income are more or less proportional.

One hypothesis is that age associated variations in average earnings parallel variations in productive capacity. Its rationalization is as follows: earnings data indicate that an individual's productive capacity reaches an age peak that occurs later in life, the more important are experience and education, and the less important physical strength and dexterity, in his occupation.[31] It seems reasonable to suppose that the greater the relative importance of the training-experience component in a given occupation the greater will be the average number of years of schooling of its practitioners. It follows that average (arithmetic mean or median) earnings will peak later among groups of persons with more (average) years of schooling than among those with less.

Whether the hypothesis of life cycles in productive capacity could be used to develop a consistent theory of the interrelation of age, earnings, education and occupation, without reference to cost of training considerations is not clear. In any case, it has never been done and at this point it seems quite unnecessary in view of the work of Mincer [10, 11]. Mincer has shown how the costs of training generate, for those with relatively more schooling (1) greater average lifetime earnings; (2) steeper average interage earning gradients until the year of peak earnings; (3) greater earnings dispersions within specific occupations or (years of schooling) groups and (4) greater interoccupational dispersion of earnings at given ages. Mincer's earlier work dealt mainly with the effects of schooling. In a later paper he extended the argument to cover on-the-job training [30].

The work of Becker, Mincer and others is based upon one particular

---

[31] Good summaries of the empirical evidence can be found in Mincer [10] and Lydall [18]. There are also some interesting results presented in Strumilin [28].

view of human capital formation. This view assumes that an individual produces human capital by investing his own time and other resources during a training or schooling period and obtaining a return during a subsequent "productive" period. Periods of on-the-job training and of part-time work while at school combine investing and producing but this complication makes no difference of principle.

So long as education or training is general,[32] the human capital produced is embodied in the individual receiving it, and he can sell its services (though not the asset itself) to whomsoever will buy. But if the training is *specific* a somewhat different situation arises. Specific training creates human capital that is jointly owned by a specific employer and the employee in whom it is embodied.[33] Either can, by severing the employment relation, unilaterally destroy the asset; therefore cooperation of both is essential to its use. The fact that the asset is jointly owned creates a field in which two-party bargaining can operate to affect the distribution of its yield between its owners.

Because of this, the decision-making machinery concerning hiring, firing and wage setting has very important effects upon the distribution of earnings. To see this, consider the hiring process for permanent jobs, especially where suitability for promotion is an important consideration. The employer invests time and money in screening applicants; successful applicants are then given more or less specific training. Subsequently, after the elapse of some time, they are again screened for promotion, for further training, etc. Hiring from outside may occur at any stage of this process, but only if suitable insiders in sufficient numbers are unavailable; *ceteris paribus* insiders are preferred to outsiders.[34]

The "promotion from within" story is an old one, but it has a substantial degree of validity. It is entirely compatible with the prevailing theory of investment in human capital once it is recognized that a substantial part of such investment takes the form of employer knowl-

[32] Becker [3, pp. 11–28] distinguishes between "general" and "specific" training. General training is that whose marginal product is the same to a large number of firms; specific training has a marginal product of zero to all employers except one. Intermediate cases can be imagined, but have not been analyzed thus far.

[33] The capital is jointly owned in the sense that its utilization requires the consent of both owners (employer and employee).

[34] The reason is simple: insiders have already been screened and given relevant specific training; outsiders have not.

edge of employee capability and probable behavior. Very often what an employer searches for via a screening process is evidence of capacity to learn and also of "good character"; i.e., evidence that the individual is unlikely to pilfer, malinger, quit or (sometimes) encourage union-ization.[35] With this knowledge as a basis, investment is made in specific training whose use provides further knowledge of capacity and character, etc.

Due to technical progress, the human capital created by a given firm may depreciate very rapidly. Yet the year to year movement in the earnings of the workers who embody it may not reflect this fact; the employer having begun to retrain them—if training was not actually continuous—as soon as their coming obsolesence became clear. That is, a company's employment policy may be one of offering employees a life-time career, including frequent "injections" of training (formal or otherwise). What the employer hires is a mixture of current labor services, L_1; detailed knowledge of the individual's capacities relevant to promotion, L_2; and a reserve of trained labor services available to meet increased demand, L_3. The demand price for L_2 and L_3 is clearly conditioned by the employer's knowledge of the worker's present capac-ity, and his estimate of the worker's capacity to learn as this has been indicated by his past performance. A satisfactory worker who has been employed for some time by employer A is clearly worth more to him than to employer B to whom he is a more or less unknown quantity requiring screening and observation. This is, of course, in addition to the difference in his value to A and B that results from specific training; the latter difference would exist even in a world of perfect information.

The relative importance of the different kinds of benefit (L_1, L_2 and L_3) that an employer expects to derive from a given act of hiring will vary both with the type of worker, and with the state of the labor market for that type of worker. Individuals hired with the thought of a permanent connection and the expectation of receiving a substantial (employer) investment in training are expected to make relatively large contributions to L_2 and L_3 and relatively small ones to L_1; those hired for only short periods and who are expected to receive little training

---

[35] "Character" is closely related to probability of quitting with resulting loss of employer investment in specific training.

are expected to make a relatively large contribution to L_1 and a small one to $L_2 + L_3$.

To oversimplify grossly, let us dichotomize the labor force into permanents and temporaries: the latter contribute only to L_1, but the former make a substantial contribution to $L_2 + L_3$ as well as to L_1. Permanents include most skilled and many semiskilled manual workers as well as white collar workers and executives; temporaries include some semiskilled and most unskilled workers. Each hiring of a permanent involves substantially larger costs for selection and initial training than is associated with the hiring of a temporary. Consequently, for equal declines in demand for current output, permanents are much less subject to layoff than temporaries.[36] Moreover, to avoid loss of their investment in selection and training, employers use indicators of a low propensity to quit (marital status, age, education, race, letters of recommendation, etc.) as important criteria for selection among applicants for permanent posts.

When a firm is confronted with a need for additional workers to fill jobs above the starting level, it has a choice between promoting those already hired (promotion from within) and hiring from outside. Even apart from important morale effects, which only strengthen our conclusion, the greater prime costs of hiring and giving (specific) training to outsiders militates against hiring them unless there are no suitable candidates within the organization. This, along with nonvested pension rights and other benefits that accumulate with seniority, explains the well-known sharp decline in interfirm mobility that comes with age.

The hiring cost bias in favor of promotion from within puts permanent workers in a situation in which an over-all increase in labor demand accelerates their progress up a job ladder. This accelerated progress of itself increases their earnings which reduces the labor market pressure on any individual firm to raise wages on particular jobs to prevent defections. However, if all firms are simultaneously confronted with an increase in labor demand some will probably have to bear the cost of

---

[36] This argument is made, with supporting evidence by Oi [33] and Becker [3, pp. 24–25]. Mincer [30, p. 69] expresses doubts as to the conclusiveness of Oi's evidence on the ground that the observed relation may be spurious because of failure to standardize for age and to distinguish between general and specific on-the-job training. The argument advanced here leads to the same conclusion as that of Oi and Becker, but Mincer's (negative) conjecture needs to be investigated.

hiring from outside at above customary starting ranks. This will create opportunities for costless (to workers) interfirm transfers for individuals part way up the job ladder thereby engendering labor market pressure on wage rates at these ranks.[37] Conversely, declines in over-all labor market demand will be manifested in slackened rates of "job progress" as well as in wage cuts or involuntary job separations or both.

The situation of temporary workers is obviously different. From the employer's point of view, a continuing employment relation with such workers is of little or no value. Hence declines in demand for output are reflected in dismissals and, depending on circumstances, wage cuts while increases in demand generate unfilled vacancies or wage rate increases. To put the matter in a slightly different way, L_1 varies more closely with demand for current output than $L_2 + L_3$. Hence over the course of a business cycle, demand for temporaries varies more than demand for permanents.[38]

For reasons already indicated, employers are much more selective of permanents than of temporaries. Also, many temporaries prefer permanent jobs from which they are, in "normal times," excluded by screening. However, when the demand for current output is strong enough, L_1 rises, relative to $L_2 + L_3$, sufficiently to induce employers to relax hiring standards on permanent jobs. That is, they find that the L_1 component on a permanent job has risen sufficiently to warrant bearing additional costs per hiring (on account of higher losses from turnover, higher failure rates at on-the-job training, etc.) in order to maintain a given rate of new hires. Relaxing hiring standards is thus a substitute for raising wage rates while maintaining a constant hiring rate.

Increasing the fraction of the labor force put through the screening and training process increases the fraction effectively eligible for permanent jobs. This is because one of the functions of hiring standards is to hold down the number of persons given a trial on a permanent

[37] Job ladders and related concepts are discussed by Kerr [34].

[38] Put into the human capital framework, our argument may be restated as follows: divide the (hourly) compensation for permanent workers into wages and "employer financed specific training." The effect of an increase in demand for permanents is reflected, in part, by extending "employer financed specific training" to individuals who would not have qualified previously; conversely for a decrease in demand.

job. Relaxing the standards, presumably, reduces the fraction of "successes" per hire, but it also increases the total number of successes.[39] This has the effect of relatively increasing the supply of permanents, and relatively decreasing that of temporaries, for a substantial interval after an initial period of labor market tightness has relaxed. This explains, in part at least, the quasi-permanent reduction in the skill differential that took place after each of the two world wars [36, pp. 269–76].

It is tempting to identify temporary workers with unskilled, and to a considerable extent I shall do so. However, this is an exaggeration that does violence to such facts as the short-term employment engagements of self-employed professionals (lawyers, physicians, accountants, etc.) who possess general training and are hired as experts for short periods.[40] It is probable that it also does violence to the wage-employment behavior of those unskilled workers who do have permanent job affiliations. Nevertheless, the overlap of the categories of temporary and unskilled is very substantial.

The relevance of the distinction between temporary and permanent workers to the size distribution of earnings arises from its common relation to risk of unemployment. For example, Mendershausen [31] found that the coefficient of concentration of incomes among "lower income" urban recipients increased from 1929–33. Mendershausen interprets this as reflecting the behavior of wage income in the face of great unemployment. Kravis [7, pp. 215–16] also seems to conclude, though hesitantly, that the bulk of the American evidence indicates that an increase in the unemployment percentage has been associated with an increase in the coefficient of concentration of urban family

[39] For example, during World War II members of minority groups, especially Negroes, were first given opportunities to acquire skills associated with permanent jobs, and many remained thereafter as part of the skilled labor force.

[40] In order to function in this manner, it is necessary—though not sufficient—that such persons be generally trained. One reader has commented that the fact of short-term engagements of self-employed professionals must be reconciled with the identification of the categories of temporary and unskilled. The reconciliation lies in the distinction between general and specific training: the common characteristic of unskilled workers and highly skilled professionals who are hired on a short-term basis is lack of *specific* training. A lawyer hired by a corporation for a particular case is hired because of his general training, and is not subjected to a lengthy indoctrination in company policies; a staff lawyer hired more or less permanently might well be given extensive specific training.

incomes. Soltow [32] found that, during the 1930's, the prolonged secular decline in the coefficients of concentration of income within a number of Norwegian cities was temporarily but sharply reversed. He (Soltow) [32, pp. 94–100] argues that this interruption of trend reflected the operation of "cyclical forces." In his paper in this volume, T. P. Schultz similarly finds a variation in income concentration with the state of the labor market due to variations in skill margins and the unemployment percentage.

Before adducing more evidence, let me relate that already cited to the dichotomy between permanent and temporary workers. What is alleged is that the relatively low hourly wage earners experience proportionately greater increases in unemployment when the over-all unemployment percentage, U, increases, and conversely when U decreases. Thus, declines in U reduce the annual earnings of those who receive initially low annual earnings relative to those initially earning more, thereby increasing the coefficient of concentration *within the lower section* of the earnings distribution; the converse occurs when U increases.

This relation of changes in U to changes in earnings concentration will arise if among various categories of wage earners there is (1) an inverse correlation between initial size of average straight time hourly wage rates and the algebraic size of the percentage change in U and (2) the percentage change in relative hourly wage rates does not have "too strong" a positive correlation with the initial levels.

First consider (2). It is widely agreed that in periods where major variations in U occurred (e.g., during the two world wars and major depressions such as 1920–21 and 1929–33) unskilled wage rates rose, in percentage terms, relative to skilled rates when U declined, but declined relative to skilled rates when U rose. In more moderate fluctuations, such as occur during the ordinary short business cycle, the same pattern of association (as during major swings) between changes in U and changes in skill differentials is observed but it is not quite so sharply defined, nor is the agreement on its pattern quite so general.[41] But though there may be doubt as to whether there is a cyclical pattern

[41] This has been discussed by a number of authors. A reasonably complete bibliography is in Reder [36, especially p. 260, fn. 1], a good recent reference is Soltow [32, pp. 131–35, especially Exhibit 78].

of fluctuation in skill differentials of the above kind, no one has suggested that the actual pattern is the reverse; i.e., it has not been suggested that the skill differential expands in prosperity and contracts in depression. And it is obvious that if (1) is satisfied and the percentage change in wage rates is uncorrelated with initial levels, the aforementioned relation between changes in U and changes in the coefficient of concentration of earnings will hold.[42]

Now let us consider (1). (1) reflects the hypothesis that fluctuations in U impinge more heavily on temporaries than on permanents and assumes that temporaries generally earn lower wage rates than permanents. Temporaries are disproportionately found in such occupational categories as (nonfarm) laborers and semiskilled operatives; they are also found disproportionately among young workers (14–19 years of age), especially recent labor market entrants, and among nonwhites. Examination of the behavior of unemployment percentages among these groups (in which temporaries are heavily represented) reveals that they move consistently in the same direction as U for the labor force as a whole, but with a greater amplitude.[43]

This pattern of associated fluctuations in relative wage rates, relative unemployment percentages and variations in the concentration of the size distribution of earnings reflects the special role of employer knowledge in the formation of human capital. Employers invest in the specific training only of individuals with certain characteristics, though the criteria for selection vary with the state of the labor market. That is, on-the-job training is offered to some individuals more readily than to others; the bases of selection of trainees are, for example, education,

[42] Lydall [18] shows that changes in wage differentials are, in general, biased as estimates of changes in over-all wage concentration. So far as it goes, his argument is correct; however, for the present purpose, it is sufficient that wage concentration should move in the same direction as the skill differential.

[43] For example, see *Manpower Report of the President,* March 1966, Appendix Tables A-12, A-13 and A-16, pp. 167–70. In an earlier paper [35], I discussed this same issue in a slightly different context and presented some relevant data.

The differential behavior of unemployment percentages of whites and nonwhites has been analyzed by Gilman [37]. Gilman's data show that the unemployment percentage of nonwhites has a greater proportional variability, over the course of the business cycle, than that of whites. However, this difference is due largely to the relatively greater numbers of nonwhites in unskilled manual occupations. When the data are standardized for occupation, the aforementioned cyclical differential disappears. I interpret this as support for the hypothesis advanced in the text.

performance on written tests, absence of a police record, absence of physical defects, etc.

The basis of selection generally used tends adversely to affect the employment opportunities of members of racial or ethnic minorities. However these bases need not imply a "taste for discrimination," though such a taste probably exists. A policy of rejecting applicants for permanent jobs on the basis of race, ethnic origin or obvious marks of low social status may be entirely consistent with a policy of maximizing some function of (private) pecuniary income. Where these socio-ethnic characteristics are strongly associated with high quit rates, inferior education, poorer health, etc., and where it is easier—and cheaper—to identify individuals by these personal characteristics rather than by the behavioral propensities they are presumed to betoken, the cheapest way to select employees may be on the basis of those readily ascertained social characteristics. But the consequence of such economic rationality can be racial, ethnic and (social) class discrimination in hiring.[44]

In terms of human capital formation this means that, from the *private* point of view, an employer's favorable opinion of an individual (as a prospective employee) is a valuable asset, having labor market effect much like training and education. Because favorable employer opinions result from characteristics attributed to or associated with formal education, part of the earnings differential associated with difference in years of schooling is a return to the ascription of behavioral traits (good work habits, etc.) associated with schooling.[45]

How much employers are prepared to pay for these traits depends upon how scarce they are, upon how valuable the prevailing technology

[44] Legislation requiring fair employment practices (hiring and promotion policies unrelated to race, ethnic background and sometimes age and sex of the applicants) is in the nature of a tax on the use of "economical" substitutes for detailed information about individuals.

[45] This fact may well be at the core of the problem encountered by Mincer [30, pp. 69–72] in explaining the partial regression (among various occupations) of the proportion of male wage and salary workers (employed 50–52 weeks in 1949) on median number of school years (mean incomes, proportion of workers below 25 and proportion employed in durable manufacturing or construction all held constant). The sign of the partial regression coefficient was positive rather than, as Mincer had expected, negative. The hypothesis offered here implies that employers would be more likely to offer permanent jobs to those more educated (wage rate constant) than to those less educated, which would produce a positive coefficient. Mincer, himself (p. 71), suggests this as one possible explanation of the failure of this coefficient to have the sign his own hypothesis suggested.

makes them, upon the strength of their belief that absence of the traits is irremediable (or excessively costly to remedy) by on-the-job training. That is, the earnings differentials observed among various groups classified by age, occupation, race, ethnic origin, recency of migration, etc., reflect not only the premium placed upon desired traits, per se, but also the strength of the presumed association (prejudice?) between these traits and their visible correlates. For example, if it were believed that only 10 per cent of the workers in a particular group would be satisfactory permanent employees, the relative demand for their services at a given wage rate would be lower than if the "believed percentage" were 20.

Thus measured, skill differentials reflect not only differences in *measured* human capital, but also the strength of employer preferences and beliefs that underlie hiring rules. These attitudes differ with the cultural and educational differences (relevant to capacity to absorb on-the-job training) among different segments of the population, as perceived—correctly or otherwise—by employers. In countries where such differences are large, employers tend sharply to distinguish different grades of workers from one another—to distinguish noncompeting groups—and pay relatively high premia for superiority; the reverse applies where the labor force is more nearly homogeneous.

Of course, there is a strong empirical correlation between inequality in the distribution of human capital and the cultural distances among noncompeting groups. As Becker [4] has shown, the distribution of human capital among the members of a population reflects both the distribution of ability, and differential access to resources for investment. But in a world where much human capital is created by employers as an aspect of productive activity, access to this means of creating human capital (jobs) also reflects culturally determined beliefs about relative ability to learn and desire to work of "typical members" of different ethnic groups, social backgrounds, etc.

Thus far, I have discussed the role of permanents and temporaries only in relation to *changes* in the concentration of annual earnings. However, differentials in earnings also affect the *level* of concentration. Consider the contrary to fact case of an over-all earnings distribution that consisted of two overlapping normal or log-normal distributions

(one for permanents and the other for temporaries) with differing means, but equal variances.[46] Then, if the ratio

$$PT = \frac{\text{mean annual earning of permanents}}{\text{mean annual earning of temporaries}}$$

increases and the variances of the subdistributions remain unchanged, the coefficient of concentration will also increase. Also, if the parameters and functional forms of both subdistributions remain unchanged, an increase in the percentage of the population in the distribution with the larger concentration of earnings (and a corresponding decrease in the percentage of the population in the other distribution) will cause the concentration of the over-all distribution to increase, and vice versa.

Obviously, one all-important question is whether the earnings distribution of permanents is more (or less) equally concentrated than that of temporaries. I can think of no a priori reason for an opinion on the direction of this inequality. What little evidence there is suggests that the annual earnings of the unskilled are less equally distributed than those of the skilled. Staehle [26] and Mincer [10] both found that the dispersion of annual earnings was relatively greater among the unskilled. Moreover, as Mincer pointed out [10, pp. 81–89], it is the unskilled alone who constitute an exception to the positive association of increased income inequality with increased average level of education. Also Hanoch [38] found that inequality of family income among Israeli urban wage earners was greater among new immigrants than among "veterans"; this was after standardization for family size and continent of origin. Hanoch's remarks strongly suggest that the new immigrant-veteran dichotomy closely parallels that of permanents and temporaries.[47]

The evidence presented by Kravis [7, pp. 194–201] is generally consistent with that of Staehle, Mincer and Hanoch. He found that among U. S. urban consumer units, headed by blue collar workers, income was most unequally distributed among those headed by unskilled workers, less unequally distributed among those headed by semi-skilled workers and least among the skilled [7, pp. 195–96]. How-

---

[46] Such a case could, depending upon the distance between the means and the relative numbers in the two distributions, easily be bimodal. However, this is of no significance for the present expository purpose.

[47] Hanoch [38, pp. 115–16].

ever he also found that among white consumer units, incomes for laborers and service workers (in the north and in the south separately) [7, p. 200] were more unequally distributed among white consumer units than among nonwhite units. This last finding is contrary to the implications of the other evidence cited in this paragraph and in the preceding one, if these findings are all taken to bear upon the difference in the concentration of annual earnings among permanent workers vis-a-vis temporary workers.[48]

What appears to underlie the main drift of the cited findings of Staehle, Mincer, Kravis and Hanoch is the belief that unskilled and new immigrants are more likely to be unemployed or intermittently out of the labor force than others. This view is given support by Miller [29, pp. 58–63] who found that concentration of earnings among occupational groups was greater, the greater their irregularity of employment. That is, the effect of irregular employment is to increase the concentration and presumably (positive) skewness of the annual earnings distribution among a given occupational group. However, this effect appears to be important only at the lower end of the occupational ladder where unemployment is a major phenomenon.

At the upper end of the occupational ladder, the effect of investment in human capital (as already noted) is to increase earnings variance and skewness within occupations as the average level of education in an occupation rises, independently of interoccupational differences in unemployment percentages. Thus it may be that the permanent-temporary dichotomy is appropriate to explaining size distributional phenomena at the lower end of the occupational ladder, but the human capital (or possibly some other approach) is relevant to the upper end.

Probably the same factors that generate differential employment opportunities as between temporaries and permanents also create dif-

[48] It has been argued by a critic that comparing inequality of income distribution among different occupational groups involves serious ambiguity unless one defines the groups so as to permit intragroup differences in earning capacity and to indicate how much of the intergroup differences in inequality reflect differing degrees of intragroup heterogeneity of earning power. The definitions of occupational groups used here are simply census definitions, and the only point being made is that the incidence of unemployment makes a greater contribution to income dispersion among unskilled temporaries than among other groups of workers. Other factors, including intragroup heterogeneity of earning power among fully employed workers, may also contribute to the observed intergroup differences in inequality.

ferences of employment opportunities among categories of workers within the permanent group. However, if this is true, its effect on the size distribution is so muted as to be indistinguishable due to the scanty data, crude tools of measurement and the brief contemporary time period to which the data refer. The last point should be emphasized. It is entirely possible that at an earlier time, sharp differences in individual earnings associated with social origin might have arisen through channels other than irregularity of employment or differential access to training, e.g., overt nepotism. However, these channels do not seem to have had any appreciable effect on available records of income distribution.

A final word of caution; the argument advanced concerning permanents and temporaries applies, without qualification, only where family heads are the sole earners. To the not inconsiderable extent that secondary workers are in the labor force, there are numerous complications discussed in the next section.[49]

IV. The Size Distribution of Family Income

As will become speedily apparent, this section is in the nature of a brief appendix to the body of this paper. While the topic is important, relatively little theoretical work has been done on it. To the extent that the size distribution among families differs from that among individuals, and in significant respects it does, discussion has been ad hoc and empirical; the remarks were often judicious but rarely related to any sort of theoretical framework.

I approach this problem as follows: assume the earning power of each individual and his property to be given. All property is assumed to be allocated to one individual or another without remainder, and with no joint ownership. Families are formed to enhance the utility of each of the various members subject to various restraints of which the most important are the rules of kinship that, in our society, require parents to live with young children and, to a limited degree, children to support aged parents.

[49] Originally, I had hoped to include a discussion of the relation of economic development to income size distribution with particular reference to the pioneering work of Simon Kuznets. However, I do not feel that what I have written does justice to the subject and, reluctantly, have decided to omit this section, hoping to return to it on another occasion.

Subject to these restraints, the formation of a family may be treated as a voluntary coalition of its members. The existence of such a coalition implies that there is one or more allocations of benefits among the family members such that each is better off within the family coalition than outside it. Were this not the case, he would leave.

The principal economic advantage of coalescence is economy of scale. These economies arise mainly in connection with housing costs; consequently the price of housing services (house rents, for short) is of crucial importance in this connection. In advanced economies at the present time, domiciliary separation of adults from all adults other than spouse is regarded as a superior good; i.e., as the combined income of a group of adults rises, the number of marriages constant, so does the number of households (recipient units) that will be formed. As the minimum number of persons per household is one, and separation from a household is an all or none matter, there is a maximum number of households that a given biosocial family (or entire community) can form. This implies that, beyond a certain point, increases in per capita income will cease to increase the number of families that will be formed from a given population.

The effect of income level on family formation works as follows: At both ends of the age distribution, young adults and retired and semi-retired persons prefer to have their own domicile, if they can afford it. Generally, the husband-wife nucleus of the family also prefers that other adult members, especially older ones, live apart. Hence as the collective income of a potential family group rises, so will the propensity of its unmarried adult members to form separate units; call this the "separation propensity." [50] Also, the lower their average income, and the higher the average house rent, the greater is the propensity of unmarried individuals to share housing in order to save expenses.

In practice, in the process of separation, each individual takes his earnings, transfers from outside the family, and property income with him. Given that earnings are the primary source of income for most individuals, and that typically earnings are appreciably higher in the middle years than at the beginning or end of the life earning cycle, the

---

[50] The separation propensity of a household will depend not only upon its level of income, but also upon its distribution among the members. E.g., freedom from parental supervision may have the highest priority for a young adult, but parents often give it a somewhat lower rating.

effect of an increase in the separation propensity is to increase the relative number of recipient units (consisting disproportionately of the very young and those past retirement) with incomes well below the average.

To put the matter in a slightly different way, suppose that as the per capita income of a community increases, the minimum income which any individual living alone must accept (social minimum income) also rises proportionately.[51] Then, as per capita income rises, house rents constant, individuals will detach themselves from higher income households, and establish new low (near the social minimum) income units. As these individuals take with them less than an equiproportional share of the (original) household's income, the effect is to increase the fraction of all households at or near the bottom of the distribution.

Whether this will increase distributional inequality, or the reverse, depends crucially upon how inequality is measured. Almost surely it will decrease the share in total income of the lowest decile or quintile of recipient units because it creates a number of new units at or near the social minimum. But its effect upon the share of those in (say) the top decile might go either way depending upon (1) the positions of the families losing members in the community's family income hierarchy and (2) upon the share of original family income taken by departing members.

(1) If the families undergoing separation were all below the top decile of income receivers, then the income share of the top decile would be unaffected. The greater the relative number of member-losing families in a given percentile, *ceteris paribus,* the greater the decline in that percentile's income share. (2) The more nearly proportional the income share taken by the departing members to that kept by the remaining members, the greater the decline in the income share of the member-losing families. Proportionality of family income shares among family members depends mainly upon the degree of equality of earning power among family members and upon the ratio of retirement income per person to the earnings of principal earners.

The effect of what has here been termed the separation propensity has been noted by Kuznets [39], Brady [40] and others. Kuznets [39] has contended, in effect, that the operation of the propensity during the

---

[51] Either because of social legislation, the state of the labor market or both.

last two decades has contributed to the apparent reversal in the move-
ment toward greater equality among recipient-units during the preceding
quarter–half century. However, this contention has been disputed by
Miller [41]. Schultz's results (in his paper in this volume) suggests that
relative increases in the 18–24 year old component of the labor force
might have caused a widening of the income size distribution since 1945,
even if the separation propensity had been unrelated to per capita
income.[52] I.e., young persons tend to be relatively low earners, and
given that a *constant* percentage of those at any specific age from their
own households, an increase in their numbers relative to all adults will
tend to lower the income share of (say) the bottom decile of households.

The tendency for "peripheral" adults to separate from original house-
holds is only one of many channels by which interrelation among family
members' behavior may alter the distribution of family incomes, given
the distribution of individual incomes. Another channel is the effect on
age of marriage. That is, marriage involves a combining of incomes if
the wife continues working; this would tend to substitute one higher
income household for two with lower incomes. In most cases where
the family members are both young earners, this will tend to reduce
over-all distributional inequality. But a further effect arises when we
consider the possible changes in the interval between marriage and
birth of the first child at which point the mother usually withdraws
from the labor force. Yet another channel by which intrafamilial inter-
actions affect the size distribution of family income is the frequent
substitution of the wife's earnings for her husband's while the latter is
studying for an advanced degree. Still other channels can be readily
imagined.

Exploring these various channels is a matter for detailed and pains-
taking empirical investigation. The Brady-Klein paper in this volume
is an excellent example of the detailed work required. This work, while
requiring and often displaying great analytical acumen, consists mainly
of meticulous description; it does not attempt to test hypotheses. More-
over, while this work enlarges our knowledge of the income distribution
within particular segments of the population, it does not bear directly
upon the analysis of over-all distributional inequality.

Despite conceptual vagueness and ambiguity, explanation of the de-

---

[52] I referred to his discussion earlier.

gree of over-all distributional inequality is a major reason for studying income size distributions, and always has been. A major, if sometimes implicit, purpose of analyzing inequality is to facilitate ranking of alternative states of an economy, differing in one or more respects, in terms of degree of equality. Assume, for simplicity, that variance of income (properly measured) is a satisfactory measure of inequality. Then, most economists seeking to compare economic states with respect to inequality would wish to compare the variance of (mathematically) expected lifetime "family-share" incomes of the members of an individual cohort. Empirical measurements of income dispersion are imperfect proxies for this theoretical concept.[53]

At any given stage of his life, an individual's actual family-share income is his imputed share of the income of the family in which he is then located. Thus if individuals with relatively low earning power typically marry those with higher earning power, family-share incomes will be more equally distributed than individual incomes. The same will be true if parents with higher incomes tend to have more children than those with lower incomes, etc. As might be expected, the family-share incomes of individuals are somewhat different from their own life earning cycles; e.g., the life cycle of highly educated husbands in family-share income will show less fluctuation than in individual income because his low earnings in the early phase will be partially compensated by his wife's earnings which will shrink after his own climb in the post-training period.

The different life cycles in family-share incomes for typical persons of different educational statuses generates—along with many other forces —a life cycle of variances within age classes. The over-all variance of a *steady state* population will be an average of these intraage class variances.[54] In the real world, "weight shifts" due to changes in relative numbers in the different age classes will produce changes in the over-all variance (degree of inequality). The significance one attaches to such

[53] In the following remarks I abstract from the difficult and important problems of correcting for differences in the effect of family composition—especially number in family—on real income. For the present purpose, income is *money income*. This may and has led to serious errors where money and nonmoney income are substitutes. But rather than offer a superficial treatment of a difficult subject, I shall simply avoid the problem.

[54] By a steady state, I mean one in which the number in each age class, in every possible cell, is equal.

weight shifts should depend upon the inference drawn for the under-
lying change—if any—in the variance of expected lifetime family-share
incomes. If there is no change inferred—e.g., if an observed weight
shift is believed to reflect only a shift in the rate of population growth—
then an increase in observed (over-all) variance has no implications for
economic inequality, and conversely. It should be noted that changes in
life expectancy, age at birth of first child and other demographic phe-
nomena, as well as in lifetime earning variances, will affect over-all in-
equality.

But even within a steady state situation there are conceptual prob-
lems. Lifetime inequality is some sort of average of degrees of inequality
at different life cycle stages. If greater lifetime inequality parallels greater
inequality at each stage, then at least ranking of alternative economic
states, by degree of inequality, is simplified. But if there is greater in-
equality in stage 1 in state A and greater inequality in stage 2 in state B,
then comparing over-all inequality as between A and B requires weight-
ing the degrees of inequality in the two stages. To weight them equally
is possible—and implicitly what has always been done—but arbitrary.
To discount incomes so that what we analyze is the variance of the
present value of expected lifetime family-share incomes is conceivable
in principle. In practice, capital markets are imperfect and, moreover,
there is no reason why the variance of lifetime incomes under conditions
of Pareto-optimality should be especially relevant to social policy. Still
further, as Kuznets pointed out in his remarks at this conference, inter-
generational inequality (over several lifetimes) may be very important.

In short, measuring and comparing income inequality as between two
situations involves serious conceptual problems akin to those of welfare
economics. These problems will not be solved quickly.