NBER Working Paper Series

A THEORY OF THE PRODUCTION AND ALLOCATION OF EFFORT

Gary S. Becker

Working Paper No. 184

CENTER FOR ECONOMIC ANALYSIS OF HUMAN BEHAVIOR
AND SOCIAL INSTITUTIONS

National Bureau of Economic Research, Inc.
204 Junipero Serra Boulevard, Stanford, CA 94305

July 1977

This research is supported by grants to NBER from the Alfred P. Sloan Foundation and the Lilly Endowment.

Table of Contents

I. Introduction

Economists have been quite successful in recent years in using simple measures of human capital - mainly, years of schooling and the length of time spent at work - to explain differences among individuals in their earnings. For example, about 30 percent of the variance in earnings has been explained by these variables in the United States, France, Chile, Sweden, and other countries (see Mincer, Riboud, and Klevmarken). Yet notwithstanding this remarkable success of the human capital approach, still well over half of the variation in earnings remains to be explained. No doubt better measures of human capital, especially of post-school investment, and of basic ability and non-pecuniary attributes of different jobs will explain an appreciable part of the remaining variation. So far, however, efforts along these lines have had very modest success (see, for example, Taubman).

It has long been recognized that earnings depend not only on hours worked and skill, but also on the effort devoted to each working hour. This has been the source of traditional arguments against long working hours and in favor of piece rates - the physical and mental fatigue from long hours lowers the effort expended per hour, and payments that depend mainly on time spent discourage effort. However, in spite of this awareness of the importance of effort, it has not been systematically incorporated into the economists' models of choice (partial exceptions are Pencavel).

The purpose of this paper is to analyze systematically the production of effort and its allocation among different market and non-market sectors. I believe that this analysis can explain much of the variation in earnings that is not explained by human capital. Moreover, it also explains many other empirical findings, and resolves some puzzles. Why is the value placed on time used in commuting and other non-working activities significantly below a person's wage rate? Why are the wage rates and hours worked of different persons positively related, even after accounting for the effect of human capital on earnings? Why does marriage raise the wage rates of men and lower those of women, and raise the health of men more than of women? Why does an increase in income often lower health? When do firms use a piece rate system? Why was the whipping and other physical punishment of slaves not uncommon? These are among the many implications of the analysis of effort in this paper.

The next section develops the basic theoretical analysis of the production and allocation of effort by a free person. Section III applies this analysis to the value placed on time allocated to the non-market sector, the effect of hours worked on fatigue and earnings, life cycle variations in earnings and hours worked, investment in health, and the effect of marriage on the earnings and health of men and women. Section IV considers worker effort from the view point of firms, and shows how various characteristics of firms determine the wage rates offered and the effort supplied by their workers. Section V analyzes the production and allocation of effort by slaves, and derives "expropriation rates" and other implications about the treatment of slaves.

II. Basic Theory

Although the word "effort" is used throughout this paper, it cannot yet be satisfactorily measured, nor adequately defined in words. Surely "effort" includes the expenditure of physical energy, but what about mental concentration, and "pressures" from deadlines and other sources? This paper sidesteps these difficulties by defining effort analytically by its role in a model of behavior. This analytic definition does distinguish increased effort from increased other human capital, and can be tested by the many observable implications of the model, even when effort itself cannot be measured. Of course, the analysis would be more powerful if effort could be measured, and I discuss some attempts at measurement in a later section.

The source of the value placed on the effort supplied to any activity is a production process that uses effort as one of the useful inputs:

$$Q_i = Q_i(E_i,\ t_i,\ K_i)\ , \tag{1}$$

where Q_i is the output in the ith activity, E_i is the total input of effort, t_i is the labor input in hours, and K_i refers to physical capital and other inputs. If the production function were homogeneous of the first degree, and depended only on the total quantities of these inputs, as in equation (1), output could be written as

$$Q_i = q_i(e_i,\ k_i)t_i\ , \tag{2}$$

where q_i is output per man hour, $e_i = E_i/t_i$ is the effort per man hour, and $k_i = K_i/t_i$ is physical capital per man hour.

It immediately follows from the assumption of constant returns to scale that the elasticity of total output Q with respect to total effort E, or output per manhour q with respect to effort per hour e, must be less than unity.[1] If there is diminishing marginal product from increases in effort,[2] the same amount of effort would be used with each man-hour or machine. In particular, it would not be profitable to extract much effort from say the first hour of a worker and little effort from his last hour (more on this in Section IIIc).

By analogy with the assumption usually made for different kinds of labor or capital, it is tempting to assume that competitive firms are able to buy effort as well as time at a fixed price per unit of service. Such an assumption, however, presupposes that the effort of one employee can be combined with the time of another employee in arriving at the optimal quantities of effort and time to each firm. The effort of any employee cannot be separated from his time, however, because his effort can only be supplied through his time. Consequently, his earnings would be zero when either his time or his effort were zero,[3] and would depend on the interaction between the quantities of each.

A more appropriate formulation assumes that firms buy a <u>package</u> of effort and time from each employee; his payments are tied to this package rather than separately to units of time or effort. The earnings of any employee would depend on the package he supplies, as in

$$I = I(E, t), \tag{3}$$

where $I(0, t) = I(E, 0) = 0$ because nothing would be offered an employee who does not supply any time or any effort. In equilibrium, an increase in either effort or time would increase earnings partly because both are assumed to have positive marginal productivity in the activity being considered, and partly because both have positive marginal values to workers since they can be used productively at other activities.

Assume, to simplify the discussion, that each firm is indifferent to the distribution of a given total number of hours among different identical workers. Then firms would offer each potential employee an earnings function that would be proportional to his hours worked, and would also depend on his productivity and the effort he puts into each hour of work:

$$\left.\begin{array}{l} I = w(e)t, \\ \text{where } w(0) = 0, \text{ and } \frac{dw}{de} = w' > 0 \end{array}\right\} \qquad (4)$$

The simplest function that incorporates these properties is

$$I = \alpha e^{\sigma}t = \alpha E^{\sigma}t^{1-\sigma} = \alpha t', \qquad (5)$$

with $t' = e^{\sigma}t$, where α is a constant greater than zero that depends on a worker's productivity, and σ is a constant that measures the elasticity of earnings with respect to a change in effort per hour.

Clearly, $0 < \sigma < 1$; otherwise, an increase in hours worked (or in effort per hour) reduces earnings if total effort (or hours) are held constant. This is inconsistent with a positive marginal product of time or effort (see equation 1), and also with a positive value of time or effort to workers at non-market activities. I have introduced the term t' to show that equation (5) can be interpreted as saying that earnings are proportional to the "effective" quantity of time t', which depends on effort per hour as well as number of hours.

The aim of firms is to choose values of σ and α that maximize their income, subject to their production functions, and to the effect of σ and α on the effort supplied per hour. Workers, on the other hand, maximize their utility by selecting the appropriate allocation of their time and effort to different activities, subject to their household production functions, and to the α's and σ's offered them by different firms. An analysis of the decisions by firms and of market equilibrium is postponed until Section IV.

For each person there is a function determined by market forces that gives his maximum earnings at different levels of hours worked and effort per hour:

$$\left. \begin{aligned} I = w(e_m)t_m &= \alpha_m e_m^{\sigma_m} t_m \\ &= \alpha_m E^{\sigma_m} t_m^{1-\sigma_m} = \alpha_m t'_m \end{aligned} \right\} \qquad (6)$$

where the subscript "m" refers to time, effort, etc. in the market sector. I say "maximum" earnings because the amount that can be earned from a given effort may be different in different firms, and workers are assumed to choose the firm with the maximum earnings. Since one firm may be best at one level of effort while other firms would be best at other levels, the earnings function in equation (6) would be constructed from the offer functions of the firms that are best at each level of effort.[4]

Clearly, $\alpha_m > 0$, and $0 < \sigma_m < 1$ for the same reasons that the corresponding coefficients in the offer functions of firms (see equation (5)) satisfy these conditions. However, if the offer coefficients of a firm were constant, these market coefficients would not be, and conversely, the market coefficients could be constant even if the offer coefficients were not. To simplify the presentation, I assume that the market coefficients α_m and σ_m are constant.

Time and effort not used to earn income can be used productively in the non-market household sector. To be specific, each household is assumed to self-produce a set of commodities with its purchases of market goods and services, and with its time and effort not spent in the market sector:

$$Z_i = Z_i(x_i, t_i, E_i), \; i = 1 \ldots n \qquad (7)$$

where x_i, t_i, and E_i are the market goods and services, time, and effort respectively used to produce the ith commodity. Following the treatment of earnings, I assume that time and effort combine to produce "effective" time input also in the non-market sector. That is, the production function for

Z_i can be rewritten as

$$Z_i = Z_i(x_i, t_i') ,$$

where
$$\left.\begin{aligned} t_i' &= w_i(e_i)t_i \\ &= \alpha_i e_i^{\sigma_i} t_i = \alpha_i E_i^{\sigma_i} t_i^{1-\sigma_i} . \end{aligned}\right\} \tag{8}$$

Clearly, $0 < \sigma_i < 1$, for otherwise either time or total effort would have a zero or negative marginal product in the production of Z_i. If increases in total effort had diminishing effects on output, the same amount of effort would be used with each unit of time, and e_i and t_i would be the only relevant decision variables for effort and time.

The sum of the time spent on each commodity and the time spent at work must equal the fixed total time available:

$$\sum_{i=1}^{n} t_i + t_m = t_h + t_m = t, \tag{9}$$

where t_h is the total time spent in the non-market sector.

The total supply of effort at the disposal of a person during any period of time can be altered by reallocation of his supply across time periods, and by the production of additional supplies. For the present, however, I assume a fixed supply of effort during any time period that can be divided between the market sector and different household activities, as in

$$\sum_{i=1}^{n} E_i + E_m = E, \tag{10}$$

where E is the total effort available. Since effort per hour is a more fundamental choice variable, equation (10) will be written as

$$\sum^{n} e_i t_i + e_m t_m = \bar{e}t = E, \tag{11}$$

where $\bar{e}$ is the average effort spent on each of the t available hours. One novelty of (11) is that the choice variables e_j and t_j enter multiplicatively rather than linearly, as do the choice variables in many household problems. This implies that the allocation of time directly "interacts"[5] with the allocation of effort.

The total expenditure on market goods and services must equal money income:

$$\sum p_i x_i = w_m t_m + v = I + v = Y, \tag{12}$$

where Y is money income, and v is the income from transfer payments, property, and other sources not directly related to earnings. Money income is not simply given to any household since it is affected by the time and effort allocated to the market sector.

Each household[6] maximizes a utility function of the self-produced commodities

$$U = U(Z_1, \dots Z_n), \tag{13}$$

subject to the production functions given by equation (8), and the constraints on time, effort and goods given by equations (9), (11) and (12). If utility were considered a derived function of goods and effective time, with marginal utilities given by

$$\frac{\partial U}{\partial t_i} = \frac{\partial U}{\partial Z_i} \cdot \frac{\partial Z_i}{\partial t_i} \; ; \; \frac{\partial U}{\partial x_i} = \frac{\partial U}{\partial Z_i} \; \frac{\partial Z_i}{\partial x_i} \;,$$

the maximization would be with respect to the decision variables x_i, t_i and t_m, and e_i and e_m.

By maximizing a Lagrangian expression that includes the utility function and the constraints on goods, time and effort, the following equilibrium conditions are easily derived:

$$\left.\begin{aligned}\frac{\partial U}{\partial x_i} &\equiv U_{x_i} = \tau p_{x_i}\\ \frac{\partial U}{\partial t_i'} w_i = \frac{\partial U}{\partial t_i} &= U_{t_i} = \mu + \varepsilon e_i\\ \tau W_m &= \mu + \varepsilon e_m\\ \frac{\partial U}{\partial t_i'}\left[t_i \frac{dw_i}{de_i}\right] = \frac{\partial U}{\partial e_i} &= U_{e_i} = \varepsilon t_i\\ \tau t_m \frac{dw_m}{de_m} &= \varepsilon t_m ,\end{aligned}\right\} \qquad (14)$$

where τ, μ, and ε are the marginal utilities of income, time, and effort respectively.

The interpretation of these conditions is straight forward. The second and third set say that, in equilibrium, the marginal utility of an additional hour spent at a particular activity equals the sum of the opportunity cost of this hour in time (μ) and in effort (εe_j). An additional hour of time has an effort as well as a time cost because a certain amount of effort, e_j, would be combined with this hour. The fourth and fifth set simply say that the marginal utility from an increase in the effort per hour must equal the opportunity cost of this effort, εt_j.

Upon substituting the third into the second set of conditions, one obtains

$$\left.\begin{aligned}U_{t_i} &= \tau[w_m - \frac{\varepsilon}{\tau}(e_m - e_i)]\\ &= \tau\hat{w}_i ,\end{aligned}\right\} \qquad (15)$$

where $\hat{w}$ is the shadow price or cost of using an additional hour at the ith household activity. The shadow price of household time does not simply equal the wage rate, but is less by a term that depends on the monetary value of an additional unit of effort, $\frac{\varepsilon}{\tau}$, and the difference between the effort per hour in market and household activities. Another expression for the shadow price of household time can be obtained by combining the last two sets of conditions, and using the relation between $U_{t_i'}$ and U_{t_i} :

$$U_{t_i} = \frac{\tau w_m' \cdot w_i}{w_i'} = \frac{\tau w_m (1 - \sigma_m)}{(1 - \sigma_i)} , \tag{16}$$

$$= \tau \hat{w}_i ,$$

where $w_m' = \frac{dw_m}{de_m}$, and $w_i' = \frac{dw_i}{de_i}$.

As a necessary condition for utility maximization, each household would select the combinations of goods and effective time that minimize the cost of producing different commodities. Effective time can be substituted for goods either by reallocating time or by reallocating effort from the market to the household sector. Household production costs are minimized when the marginal rate of substitution between goods and effective time equals the cost of converting either time or effort into market goods. This is readily shown by combining the first and last two sets of equilibrium conditions, and the first three sets to get, respectively,

$$\frac{\partial U}{\partial x_i} \Big/ \frac{\partial U}{\partial t_i'} = \frac{\partial Z_i}{\partial x_i} \Big/ \frac{\partial Z_i}{\partial t_i'} = \frac{p_i w_i'}{w_m'} , \tag{17}$$

$$\frac{\partial U}{\partial x_i} \Big/ \frac{\partial U}{\partial t_i'} = \frac{\partial Z_i}{\partial x_i} \Big/ \frac{\partial Z_i}{\partial t_i'} = \frac{p_i w_i}{w_m - \frac{\varepsilon}{\tau} (e_m - e_i)} , \tag{18}$$

where the right hand side of (17) gives, for the ith commodity, the ratio of the costs of using goods and effort, and the right hand side of (18) gives the ratio of the costs of using goods and time.

It has been useful when analyzing the allocation of time and goods to combine the separate time and goods constraints into a single full income constraint (see Becker, 1965). The time, effort and goods constraints given by equations (9), (11) and (12) can also be combined into a single full income constraint that shows how the maximum money income achievable is spent by using goods, time, and effort in household production:

$$\left.\begin{aligned} \sum p_i x_i + F(w_m, E_h, t_h) &= w_m\left(\frac{E}{t}\right)t + v = S \\ &= w_m(\bar{e})t + v = S \end{aligned}\right\} \tag{19}$$

where S is full income, and F is foregone earnings.

Money income is maximized when all the effort and time is allocated to the market sector because the supply of effort is assumed to be independent of its allocation, and the elasticity of earnings with respect to effort per hour is less than unity. Full income depends on four parameters: "property" income v, the wage rate function w_m, the available time t, and the supply of effort per unit of time, $\bar{e}$. Foregone earnings depends on the time and effort spent in the household, t_h and E_h, and on the wage rate function. Utility can be maximized subject only to the full income constraint; the following equilibrium conditions are readily derived:

$$\left.\begin{aligned} U_{x_i} &= \tau p_{x_i} \\ \frac{\partial U}{\partial t_i}\Big|E_i \text{ and } x_i \text{ held constant} &= \tau\frac{\partial F}{\partial t_h} = \tau w_m(1-\sigma_m) \\ \frac{\partial U}{\partial E_i}\Big|t_i \text{ and } x_i \text{ held constant} &= \tau\frac{\partial F}{\partial E_h} = \tau w_m' \end{aligned}\right\} \tag{20}$$

The last condition is a variation on the fourth and fifth conditions in equation (14), and the second one is a variation on the second and third conditions in that equation.

The right hand side of equation (15) equals that of (16) because they are different ways to state the cost of using an additional unit of time in the ith commodity:

$$U_{t_i} = w_m \frac{(1-\sigma_m)}{(1-\sigma_i)} = w_m[1 - \frac{\varepsilon}{\tau w_m}(e_m - e_i)] \qquad (21)$$

Since the dollar value of an additional unit of effort, ε/τ, equals the rate of increase in the wage rate as effort increases, w'_m - see the last condition in equation (14) -, we can derive from (21) a simple and fundamental relation on the equilibrium ratio of the effort per hour supplied to any two activities:

$$\frac{e_m}{e_i} = \frac{\sigma_m(1-\sigma_i)}{\sigma_i(1-\sigma_m)}, \text{ all } i \qquad (22)$$

or

$$\frac{e_j}{e_i} = \frac{\sigma_j(1-\sigma_i)}{\sigma_i(1-\sigma_j)} \quad \text{all } i, j . \qquad (23)$$

The ratio of the effort per hour in any two activities depends only on the elasticities of their effective time functions, and would be constant as long as these elasticities were constant, regardless of the elasticities in other activities, the utility function, other characteristics of production functions, changes in full income, and the equilibrium allocation of time or goods. For example, a change in the elasticity in a given activity would tend to change the effort per hour in different activities, but would not change the ratio of the effort per hour in other activities if their elasticities were unchanged. This invariance property is of considerable value in simplifying the analysis of the effects of changes in different parameters on the allocation of a household's time and effort.

III. Applications

a. Effort Elasticities

Equation (23) clearly shows that the effort per hour would be higher in one activity than in another if, and only if, its elasticity of effective time with respect to effort were higher. A few things can be surmised about the ordering of different elasticities. For example, the amount of sleep is obviously closely dependent on the time but not on the effort[7] devoted to sleep. Similarly for listening to the radio, reading a book, and many other leisure activities. On the other hand, the earnings from most jobs and the output from household activities like the care of small children are much more dependent on the input of mental and physical effort.

A major implication of this brief discussion, that the elasticity of earnings with respect to effort is usually larger than is the effort elasticity in most household activities, is supported by the empirical evidence on the value of time. According to equation (16), the time used in a particular household activity would be valued at less than the wage rate if, and only if, the effort elasticity of the wage rate exceeded the elasticity of effective time at this activity. Available estimates of the value of time are usually much below the relevant wage rate, although practically all these estimates are for time used in transportation.[8]

Accordingly, I make the important assumption in the rest of this paper that elasticities of wage rates with respect to effort generally exceed elasticities of effective time with respect to effort in the household sector, not only on the average but also at the margin in the sense that an exogenous increase in hours worked and corresponding reduction in hours spent in the household would increase the demand for effort. This assumption does not deny that these elasticities have considerable variance within both the market and household sectors; for example, the elasticity for coal miners greatly exceeds that for night watchmen, or the elasticity for baby care greatly exceeds that for sleep. Moreover, the analysis in later sections demonstrates that these elasticities are not simply given, but are partly the result of optimizing decisions by firms and households. For example, married men, single women, healthier persons, or larger firms, in effect, choose relatively higher market elasticities than do single men, married women, less healthy persons, or smaller firms, respectively.

b. The Value of Household Time

The cost of an additional hour of time at a particular household activity usually is not just the hourly wage rate, for any saving in effort from reallocating time away from work must also be valued. If more effort goes into an hour of work, an additional hour at this activity would save effort, and its marginal cost would be the difference between the wage rate and the money value of the saving in effort. This is shown explicitly by equation (15), for $\frac{\varepsilon}{\tau}$ is the money value of an additional unit of effort, and $e_m - e_i$ is the saving in effort.

Consequently, the cost of time would not be the same for all household activities; it would be least for those using the least effort per hour. Moreover, the cost of time at a particular activity would not necessarily be the same even for persons with the same wage rate: the money value of effort and the saving in effort from using time at this activity could well be very different.

The cost of time is written in terms of the effort elasticities in equation (16). The cost at a particular activity would equal, exceed, or be less than, the wage rate as the elasticity of effective time with respect to effort at this activity equalled, exceeded, or was less than the elasticity of the wage rate with respect to effort. Note that the cost of time can exceed the wage rate, and might well do so for effort-intensive activity like the care of young children, or do-it-yourself house painting. The discussion in the previous section suggests, however, that the cost of time would be less than the wage rate for most household activities because wage rates are more responsive to changes in effort than are household activities.

A decrease in the wage rate or an increase in its elasticity with respect to effort would lower the cost of time at all household activities by the same percentage. Therefore, the ratio of the cost at any two activities is independent of both the wage rate and its elasticity, and depends only on the elasticities at these activities:

$$\frac{U_{t_i}}{U_{t_j}} = \frac{\tau\hat{w}_i}{\tau\hat{w}_j} = \frac{1 - \sigma_j}{1 - \sigma_i} \qquad (24)$$

A given percentage decline in (one minus) the elasticity at the jth activity relative to that at the ith activity would reduce the relative cost of time at the ith activity by the same percentage.

Practically all the empirical estimates of the real value of time deal with transportation - such as commuting or airline travel -, and these estimates almost invariably are below the wage rate. Indeed, they have usually been less than 50 percent of hourly earnings.[9] Such "low" estimates have been considered troublesome,[10] and have been explained in various unsatisfactory ways, including disutility from work (Gronau), utility from time spent commuting and in other kinds of travel,[11] and discrepancies between marginal and average hourly earnings (Becker, 1965).

Once the allocation and evaluation of effort is included in the analysis, however, these estimates no longer seem particularly "low" nor troublesome, and can be readily explained without assuming disutility from work, or a difference between marginal and average wage rates. A value of time equal to say 50 percent of the wage rate implies, first of all, that the elasticity of earnings with respect to effort exceeds 1/2,[12] a plausible restriction since effort is presumably more important than time in the earnings function.[13] Moreover, if this elasticity were between .6 and .8, the implied elasticity of effective travel time with respect to effort would be between .2 and .6.[14] Consequently, neither implausibly high values of the elasticity of earnings nor implausibly low values of the elasticity of effective travel time are implied by shadow costs of travel time equal to only 50 percent of the wage rate.

Gronau's finding that the value of travel time is higher for business than for personal trips is readily explained if business travel is more like work and has a higher effort elasticity than personal travel. Our analysis can also explain why persons with higher hourly earnings commute and generally travel more comfortably: they want to economize on the effort spent in traveling because their effort is more productive at work.[15] An increase in hourly earnings would not change the ratio of the value placed on traveltime to hourly earnings[16] if the elasticities of earnings and travel time with respect to effort did not change.[17]

More generally, our analysis implies that the shadow price of the time spent not only on travel but on any other household activity as well is different for different persons. In particular, the shadow price would tend to be a lower fraction of the hourly earnings of married men, single women, middle-aged men, healthier persons, and persons with higher earnings (see the discussion in later sections). I already pointed out that the shadow cost of the time used in different activities of the same person also varies greatly; for example, a woman might value the time spent caring for her young children at much higher than her wage rate, and the time spent watching television with her husband at much lower than her wage rate.

This analysis not only explains the findings on the value of time spent in travelling, but is also relevant in estimating the shadow cost of time in child care and other household activities.[18] The allocation of effort has been ignored, and the cost of all the household time of any person has generally been assumed to equal his wage rate. If the allocation of effort has an important effect on the cost of time, its neglect could have significantly biased estimates of the cost of time. For example, women with market wage rates equal to w will care for their children or clean their houses rather than participate in the labor force as long as the marginal product of their time in child care and house cleaning exceed aw and bw respectively, where aw and bw are the shadow costs of their time in child care and house cleaning. Since my analysis suggests that b differs from a, then $\frac{bw_b}{aw_a} \neq \frac{w_b}{w_a}$; hence the methods used will not estimate correctly the ratio of the marginal products of their time spent in child care and house cleaning.

c. Hours of Work

The traditional analysis of the effect of hours of work on the effort devoted to each hour argues that fatigue, both mental and physical, sets in beyond a moderate number of hours, and thereafter grows increasingly rapidly as hours continue to increase. Eventually, effort per hour declines more rapidly than hours increase and total productivity itself begins to decline. To take a good example of this analysis, Denison explicitly assumes, in discussing the contribution of declining hours to economic growth in the United States, that when hours worked increase beyond 43 per week, the productivity of each hour declines by 30 percent, and beyond 53 hours per week (for men) total productivity itself declines.[19]

The traditional analysis is frequently justified by alleged physiological findings that the energy and other effort used up in basal metabolism, sleep, dressing, and other basic human functions is fixed, or at least must remain above a given minimum level. If so, the effort available for each working hour must eventually decline since only a limited quantity can be withdrawn from other activities. A related argument (Freudenberger and Cummins) assumes that a fixed amount of effort must be spent on each working and each household hour. If the total stock of effort is fixed and binding, and if work uses more effort than non-work, the number of effective working hours would have to be less than the total number of hours available; stated differently, beyond a certain point, additional working hours would have no effect on total output.

Although the basic human functions use up a certain amount of effort, most persons in the Western world spend their non-working time on much more than the basic functions, and have considerable leeway in their mix of non-work activities. Moreover, the effort used up on the basic functions is not rigidly fixed, but can be altered through changes in body weight, amount of rest, and in other ways (see Freudenberger and Cummins). Therefore, it would be technologically feasible over a wide range of hours worked for effort per hour to rise as hours worked rose.

Consequently, I assume that the relation between effort per hour and hours worked is a matter of choice, not of technology, although I recognize that choices are conditional on physiological and other constraints. Effort per hour declines as hours worked increases only when utility is greater with a decline than with an (technologically feasible) increase. Moreover, the relation between effort and hours could be positive under some conditions and negative under others; there is no single "the" relation, not only in magnitude but also in sign.

Consider equation (22) again, which shows that the ratio of the hourly effort at work to the hourly effort at any household activity is a simple function of their effort elasticities. A change in property income, human capital, or other variables that did not change the effort elasticities would have no effect on the ratio of the hourly effort in different activities. The hourly effort in any activity could increase or decrease, but it would change by the same percentage in all activities. A constant ratio is an implication of utility maximization and the other attributes of our model; that is, it is a theorem rather than an assumption about the technology relating earnings and output to effort. Moreover, a constant ratio should not be confused with a constant level of hourly effort in different activities.[20]

A change in property income, the stock of effort, or other variables would tend, among other things, to change hours worked. If effort elasticities did not change, if the elasticity at work exceeded the marginal elasticity in the household sector (see the discussion in Section IIa), and if the stock of effort were unchanged, then a change in hours worked would be associated with a change in the opposite direction in the effort spent on each working hour.[21] A change in hours worked would also be associated with an opposite change in the effort spent on each household hour[22]: both say decline as hours worked increase because work is more effort-intensive.

The elasticity of hourly effort with respect to hours worked would be a constant less than one (in absolute value) - total effort spent on work must rise as hours of work rose - that would be larger, the larger was the elasticity of earnings with respect to effort relative to the elasticity in the household sector.[23] Consequently, since the elasticity of hourly earnings with respect to hourly effort is less than one, the elasticity of hourly earnings with hours worked would be negative and of moderate size: using values of σ_m and σ_h that are consistent with the findings on the shadow price of travel time (see Section IIb), we can estimate the elasticity of hourly earnings with respect to hours worked at -1/2, and that of total earnings at +1/2.[24]

Although hourly earnings and hours worked were apparently negatively related 60 years ago in the United States,[25] they now appear to be positively related.[26] Hourly effort and hours worked would be positively related if hours worked differed mainly because the stock of effort differed, for hours worked and effort per hour would then vary in the same direction.[27]

d. Property Income, Human Capital, and Effort Elasticities

A change in property income initially does not change the cost of time or effort, and has only an income effect, which reduces hours worked and increases the consumption of market goods. Since a reduction in hours worked increases hourly effort and earnings, wealthy persons have higher hourly earnings partly because wealth induces an increase in hourly earnings. Although effort per hour would increase with wealth, the total effort spent on work would decrease, so this analysis at least partly supports the common view that wealthy persons usually do not work as "hard" as poor persons. The increase in hourly earnings induces a substitution toward longer hours, but since this substitution effect is created by the wealth effect, hours of work must be reduced. However, the net decline in hours worked would be less than the decline predicted from an analysis that ignores the effect of wealth on effort per hour.

Define a "neutral" or "general" increase in human capital as an increase that raises the efficiency of time in all activities by the same percentage; that is, the level parameters α_m and α_l in equations (6) and (8) are raised by the same percentage. If the increase in human capital were exogenous, full income would also increase, but the relative cost of both time and effort at any two activities would be unchanged (by the neutrality assumption). The income effect would reduce hours worked if relatively time-intensive commodities had income elasticities that tended to be above unity, and if property income increases at the same rate as human capital. A reduction in hours worked would in turn induce an increase in hourly effort. Although effort would change by the same percentage in all activities, hourly effectiveness or "earnings" would change by different percentages, even with a neutral change in human capital. In particular, if hours worked declined, the increase in hourly earnings at work would exceed the increase at household activities. Consequently, even neutral human capital would appear to be market biased in the sense that the effectiveness of market time would appear to increase more than the effectiveness of household time.

An increase in human capital that is biased toward work - human capital that raises α_m by a greater percentage than the household α_i - would induce a substitution toward both time and effort at work because the cost of using them in the household would be raised. Therefore, if any increase in real income were compensated, more hours and effort would be devoted to work. Since, however, effort per hour would fall because of the rise in hours worked, hourly earnings would rise by a smaller percentage than α_m would.

Some jobs offer wage functions that are relatively more responsive to change in effort, while other offers are relatively more responsive to changes in effective time.[28] Put in terms of the wage function in equation (6), some have larger σ_m, the elasticity of wages with respect to effort, while others have larger α_m, the compensation per unit of effective time. Since changes in σ_m and α_m have different effects on the supply of hours and effort, even jobs that are equally attractive could have different combinations of hours and effort.

Consider, for example, two jobs that are equally attractive to a given class of workers; that is, provide the same utility to members of this class. If the difference in their σ_m were small, the difference in their α_m must equalize hourly earnings in both jobs when hourly effort in one job equalled the optimal level of hourly effort in the other.[29] These persons would use less effort in the household and more at work when employed in the job with the larger σ_m because the marginal cost of using effort in the household would then be greater.[30] On the other hand, they would use less time at work and more time in the household even though the wage rate is held constant, because an increase in σ_m increases the saving in effort from substituting household for working time.[31]

It is not surprising that more effort and less time would be devoted to the job that rewards effort better and time worse.[32] Note also that workers would supply fewer hours in the job that pays the higher hourly earnings not because the increase in earnings induces a reduction in hours, nor even primarily because an increase in hours induces a reduction in hourly earnings, but because the different jobs offer different marginal costs of using time and effort in the household.

e. The Stock of Effort and Health

The stock of effort possessed by different persons varies greatly, not only in dimensions like mental and physical energy,[33] but also in "ambition" and motivation. Effort is a part of the total stock of human capital, and differences in the stock of effort, like differences in other human capital, are the result either of differences in endowments - differences in genes - or differences in investment. An increase in the endowed stock of effort raises real income because a larger output of commodities could be produced from given wage and household production functions.

The effects of a change in the stock of effort are not neutral between the market and household sectors, but are biased toward the market essentially because the market is more receptive to increased effort than are household activities. Although equation (22) guarantees that an increase in "effort capital" would increase the equilibrium effort per hour by the same percentage in all uses, the productivity of an hour of working time would increase by a larger percentage than that of the average hour spent in the household because the effort elasticity at work is larger than the average household elasticity (see Section IIIa).

That is, persons with greater effort (like Gladstone in footnote 33) would excel especially at work, in the sense that the productivity of their time would be especially large at work.

Moreover, persons with greater effort would be oriented toward work also in the sense that they would tend to work longer hours. Since their time is relatively more productive at work than at household activities, they would be induced to reallocate time away from the household and toward work. The income effect from greater effort capital may offset this substitution effect, but complete offset is unlikely since even the sign of the income effect is not clear a priori.[34] Consequently, hours of work and hourly earnings (and effort) would be positively, not negatively, related among persons with different stocks of effort.

If different jobs compensated effort and time in different ways, persons with more effort capital would be attracted to jobs compensating effort relatively well - jobs with relatively large σ_m - and persons with less effort capital would be attracted to jobs compensating time well - jobs with relatively small σ_m. Since the previous section shows that an increase in σ_m increases the supply of effort and reduces the supply of time to work, a positive relation between σ_m and the stock of effort capital would reinforce the tendency for persons with more effort capital to allocate a larger fraction of their effort to work, and would partly offset their tendency to work longer hours.

If hours of work and the effort elasticities were unaffected, an increase in the stock of effort capital by say 10 percent would increase effort per hour also by 10 percent. Output and earnings would not increase by as much because the elasticity of output and of earnings with respect to effort are less than unity (see Section II). However, the induced increase in hours of work and/or these effort elasticities could raise output and earnings sufficiently further, so that they could increase by more than 10 percent.[35] Therefore, our analysis can explain the finding in several experimental studies that the elasticity of the output of workers in heavy industries with respect to changes in their consumption of calories - one component of the stock of effort - is apparently greater than unity (see Food and Agricultural Organization, pp. 14-15, 23-25).

An improvement in health is commonly interpreted in terms of an increased supply of energy and other kinds of effort; for example, mental or physical sickness is said to reduce the supply of mental or physical energy to different activities. I support this usuage and go even further by assuming that the stock of "health capital" of any person is proportional to his stock of effort capital.[36] The main analytical and empirical advantage of such an assumption is that extensive and specific implications about the effects of health can be derived because our model has extensive implications about the effects of effort.

Consider the well known findings that ill-health reduces both hourly earnings and hours worked.[37] If health and effort were directly related, hourly earnings would be reduced because a decrease in the stock of effort decreases the effort devoted to each working hour. Of course, ill-health would then also reduce the productivity of household time because the effort devoted to each household hour would also decrease; casual observation clearly confirms this implication as well. Ill health reduces hours worked because the household sector is less effort-intensive than work; a reduced stock of effort can be used more economically by redistributing time from work to the household. Put differently, sick time is spent at home rather than at work because home activities do not demand as much effort.[38]

This way of looking at health capital implies that it is biased toward the market sector in the sense that an improvement in health raises productivity of market time by a larger percentage than household time.[39] This bias not only explains why an improvement in health raises hours worked, but also why the elasticities of market output and earnings with respect to health may exceed unity,[40] even though these elasticities with respect to effort per hour are less than unity.

The market bias of health also implies that the incentive to invest in health is greater for persons who work longer hours. To derive this and further results, assume that time, goods and effort can be devoted to exercise, sleep, physical check-ups, relaxation, proper diet, and other activities that raise the stock of health capital. The cost of using these inputs to produce more health capital is their foregone value at other activities; an optimal amount would be spent on improving health when the cost of an additional unit of health just equals its value.

The output of effort or health is related to the input of goods, x_s, time, t_s, and effort itself, e_s,[41] by a household production function

$$E = E(x_s, t_s') + E_0 , \quad \text{with } t_s' = t_s w_s(e_s), \tag{25}$$

where E_0 is the "endowed" level of E; the level when nothing is being spent to raise effort ($x_s = t_s' = 0$). By maximizing the utility function given by equation (13) subject to time, goods, and effort constraints that include t_s, x_s, and e_s, one could derive equilibrium conditions for these inputs that are comparable to those found in equations (14) or (22). It is revealing, however, to maximize also with respect to the stock of effort, and derive the equilibrium condition

$$\frac{\varepsilon}{\tau} = \frac{\varepsilon}{\tau} t_s \frac{de_s}{dE} + p_s \frac{dx_s}{dE} + (\mu + \frac{\varepsilon}{\tau} e_s) \frac{dt_s}{dE} \tag{26}$$

The term on the left is the shadow price or money value of an additional unit of effort, and that on the right is the cost of the inputs used to produce an additional unit. By using the equilibrium conditions in (14) and (22), this condition can be written as

$$w'_m = \frac{\varepsilon}{\tau} = w'_m t_s \frac{de_s}{dE} + p_s \frac{dx_s}{dE} + w_m \frac{(1-\sigma_m)}{1-\sigma_s} \frac{dt_s}{dE} . \tag{27}$$

An increase in the shadow price of effort increases the optimal investment in effort because marginal benefits increase by more than marginal costs. Since equation (27) shows that, in equilibrium, the shadow price equals the marginal wage rate with respect to effort, an increase in the marginal wage would increase the optimal investment. The marginal wage is increased by a decline in property income because hours of work would increase, by a rise in market or non-market effort elasticities,[42] and by an increase in other kinds of human capital. Therefore, health capital would tend to be greater for persons working longer hours - again indicating that the effects of health are biased toward work -, and when other kinds of human capital are greater. The converse of these statements are also true: hours of work and the incentive to invest in other kinds of human capital[43] are greater when health capital is greater, for an increase in health raises hourly earnings, which in turn increases hours worked (see the discussion above), and increases the shadow price of investments in human capital. Although the marginal cost of these investments would also increase if time were more important than effort in the production of human capital, the increase in costs must be less than the increase in returns, so that the incentive to invest would increase.

Since an increase in certain kinds of income tends to lower health (because hours worked and thus the shadow price of effort are reduced), whereas an increase in other kinds tends to raise health (because the marginal wage with respect to effort is raised), our analysis provides the solution to a very puzzling finding;[44] namely, that higher income persons in the United States have, if anything, poorer health than lower income persons. This finding does not look puzzling when health is related to effort, for even the sign of the net effect of an increase in income on the stock of effort is ambiguous, and depends crucially on why income increased. An increase in hours worked or in earning power raises the stock of effort and health, whereas an increase in earnings for other reasons,[45] or an increase in other kinds of income lowers effort and health.

These conclusions can be tested empirically by decomposing the overall relation between health and income into separate relations for earnings and other income, and even for different determinants of earnings. Moreover, a further test can be developed by recognizing than an increase even in property income would increase the incentive of persons not in the labor force to invest in effort because the value of additional effort would increase.[46] Hence health and income should be more positively related for married than for single women, or for women on welfare than for equally poor working women.

Persons investing in health and other kinds of human capital tend to work fewer hours because investment time is drawn mainly from working time. If investments used more effort per hour than work - that is, if the effort elasticity at work were less than that at investments -, effort and earnings per hour at work would decline as additional time and effort were spent on investments. Even if investments used less effort per hour than work, earnings per hour would be less for persons investing than for others also working fewer hours if the effort per hour devoted to the consumption activities where the latter spent their added time is less than the effort per hour devoted to investing. There is in fact evidence that the hourly earnings of persons in school are less than those of persons not in school who work the same number of hours and appear otherwise to have the same characteristics (see Lazear).

The foregone cost of each hour spent investing does not simply equal the hourly earnings of investors, but depends also on the effort devoted to investments: the foregone cost is less than or exceeds these hourly earnings as the effort devoted to each hour of work exceeds or is less than the effort devoted to investments (see equation (24) for a statement in terms of the effort elasticities). Consequently, the foregone earnings of investors could be larger or smaller than the product of their hourly earnings and their hours spent investing (see also Lazear), or than the product of the fraction of working time spent investing and the total earnings of persons not investing. The correct estimate of foregone earnings is simply the difference between the total earnings of persons not investing and the total earnings of investors (suitably corrected for differences in ability, etc.)

If investment in effort were put into a life-cycle context, there would be a shadow price and marginal cost of adding to effort at each age. The stock of effort capital would vary over the life cycle, tending to rise at younger ages, and eventually falling at older ages.[47] Since the stock of effort would rise and fall along with the stock of other human capital, the increase in earnings at younger ages would be partly caused by an increased stock of effort (i.e., an improvement in health), and the decline at older ages would be partly caused by a decline in effort (i.e., a deterioration in health). Indeed, when the marginal wage rate is relatively high, the supply of effort might be augmented by "borrowing" from the effort available at other ages, which would raise the hourly wage rates further at the ages doing the "borrowing", and depress them further at the ages doing the "lending"; in extreme forms, this "borrowing" is called "overworking" or "burning oneself out".

f. The Supply of Effort Within a Family

As shown elsewhere (Becker,   ), married and other multi-person households are prevalent because single persons can increase their utility by combining into a larger household. The gains from marriage are related to the production of children, love, the division of labor between market and household activities, protection against risk, and most relevant for present purposes, the division of labor in the allocation of effort.

The gains from the division of labor in effort can be studied in isolation by considering a group of men and women who differ only in their stocks of effort, which are assumed to be distributed exactly the same way among women as among men. If the household production functions had constant returns to scale, a man and woman with the same stock of effort would not gain from marrying each other because their married household would allocate resources the same way as their single households, and would produce the same output of commodities as the combined output of their single households.[48] On the other hand, if persons with large stocks of effort married persons with small stocks, the married household would produce more than the (combined) single households because married households could take advantage of the gains from the division of labor. The basic reason is that the mate with the larger stock of effort would have a comparative advantage in the market sector because earnings respond more to a given increase in effort than does the output of household commodities (see the discussion in Section IIIa). Hence that mate would supply more time and effort to work than when single, and the other mate would supply more time and effort to the household.

I have essentially shown that a full negative assortative mating of the stocks of effort is the optimal way to combine mates because it is the most efficient way to combine them.[49] The negative sorting by effort is another exception to the general rule (proven in Becker [1973]) that positive assortative mating of traits is optimal; the exceptions occur when, as with effort, the gains from the division of labor are important. I believe that casual observation confirms that energetic persons do tend to marry relatively lethargic persons, at least when other traits are held constant.

This analysis can explain why in most marriages one mate has tended to specialize in work and the other in household activities. The explanation does not presume that mates have different earnings or household functions, or that they differ in abilities or human capital that change these functions. This division of labor would follow as long as there are significant differences in the stock of effort among men and women. However, this analysis cannot explain the sexual specialization in the division of labor - the tendency for married men to work and married women to remain at home - without assuming that, on the average, men have had much larger stocks of effort. I do not know of any plausible reasons why this should be intrinsically so (as opposed to the effects of family and social policies).

Consequently, I shall simply assume that, whatever the reasons, married men have had a comparative advantage at work. Married men devote longer hours and more effort to work than single men because the time and effort of married men in the household sector can be replaced by the time and effort of their wives. Single men can use market substitutes, but these are also available to married men. The additional substitution posibilities provided by wives are generally greater in relatively effort-intensive household activities - greater in say food preparation than in sleeping (see Gronau). Hence married men not only devote more of their effort to work than otherwise similar single men because married men working longer hours, but they also devote more effort to each our of work than would single men if they worked as many hours. Consequently, married men would have higher hourly earnings and would have an incentive to choose jobs that compensate effort relatively well. The empirical evidence supports this implication of the analysis, for married men with spouse present earn considerably more than other men - perhaps 15 percent more - even when the comparisons are standardized for hours worked, human capital, and other variables.[50]

An argument can be developed along the same lines to show that married women not only work fewer hours than otherwise similar single women, but also supply less effort per hour of work. Hence married women would have lower hourly earnings and would choose jobs that compensate time relatively well and demand less effort. It is well known that married women earn much less than single women, per hour of work as well as in total (see, e.g., Fuchs 1974), and the general impression is also that married women choose jobs that demand less effort. Therefore, the division of labor in the allocation of effort within families explains part of the earnings and occupational differential between married and single persons.

One of the more interesting findings about the family is that marriage appears to raise the health of men more than the health of women.[51] This has been something of a puzzle, and explanations have been sought in the division of labor: men gain more health from marriage because all women, although especially wives, specialize in household production, including the production of health[52] (see). Our analysis also relies on the division of labor, but through market, not household, production. The effect of marriage on health is determined by its effect on the incentives to invest in effort, which need not be the same for men and women. Since these incentives are biased toward the market, marriage would raise the health of men more than that of women because married men specialize more in market activity.

Indeed, marriage would appear to lower the value of health to women since the traditional sexual division of labor has reallocated their time from the market sector. However, the negative effect on women is attenuated by the specialization of married women in child-rearing and other effort-intensive household activities. Similarly, the positive effect on the health of men is attenuated by the decline in their marginal product of effort as more of their effort is devoted to each hour of work. Still, that marriage has different effects on the value of an increased stock of effort to men and women appears to be the most promising explanation of why marriage has different effects on the health of men and women.

IV. The Market Determination of the Response of Earnings to Effort

Section II derived some properties of the elasticity of earnings with respect to effort, for example, that it is less than unity, and other aspects of the earnings function. Subsequent sections explored the consequences of changes in this elasticity and other parameters for the supply of effort and time. This section shows how the earnings function, especially the effort elasticity, is determined by the interaction between the maximizing behavior of firms and workers in the market for labor.

Assume initially that all firms and all workers are identical. Each firm is assumed to have a Cobb-Douglas production function that has constant returns to scale in the total effort and time of employees:[53]

$$Q = \beta E^{\gamma} t^{1-\gamma} = \beta e^{\gamma} t, \tag{28}$$

where Q is output, e and t the effort per hour and total time of employees, and β and γ are positive constants with $\gamma < 1$ measuring the elasticity of output with respect to effort. The hourly earnings function offered to employees is also assumed to be Cobb-Douglas:

$$w = \alpha E^{\sigma} t^{1-\sigma} = \alpha e^{\sigma} t, \tag{29}$$

where α and σ are positive constants.

The effort per hour and number of hours supplied by each worker is partly determined by these parameters of the earnings function:

$$\begin{aligned} e &= e(\alpha, \sigma, \mu) \\ t &= t(\alpha, \sigma, \mu) \quad , \end{aligned} \tag{30}$$

where μ represents other determinants. It is shown in Section III that an increase in σ compensated by a decline in α that holds the employee's utility constan would raise σ and lower t. If the increase in σ were not compensated, the increase in real income also raises e and lowers t, and thereby reinforces the substitution effect.

Each firm has some control over the values of α and σ that it offers to potential employees. If it could choose any values of α, σ, and t without incurring any "monitoring" costs, it would choose values that maximized the difference between sales and payments to employees:

$$\pi = pQ - wt = p\beta e^{\gamma}t - \alpha e^{\sigma}t \ , \qquad (30)$$

where p is the price of the product. If firms were competitors in the product market, the optimal number of hours would be where the marginal product of hours equalled hourly earnings:

$$\frac{\partial \pi}{\partial t} = 0 = pq - w,$$

$$\text{or} \qquad (31)$$

$$p\beta e^{\gamma} = \alpha e^{\sigma} \ ,$$

where q is output per hour.

Equation (31) can also be interpreted as a condition insuring that profits are zero. If profits were positive, competitive firms with constant returns to scale would expand indefinitely; if profits were negative, they would not produce anything. In market equilibrium, the quantity demanded of the product must equal the quantity supplied, where supply depends entirely on the effort and time supplied. Since equation (31) essentially determines the ratio of α to the market price p, if supply were less than demand, both α and p would rise until the reduction in quantity demanded (through the rise in p) and the expansion in quantity supplied[54] (through the rise in α) eliminated the gap between demand and supply.

If time and effort were hired in a competitive labor market, the real hourly earnings of workers would be determined by the market forces considered in the previous paragraph, and would be given to any firm. However, each firm has the freedom to choose among the many combinations of α and σ that are equally attractive to workers. A firm would choose the combination that maximized its income. By differentiating total income with respect to σ, one derives a necessary condition to maximize income:

$$\frac{\partial \pi}{\partial \sigma}^{55} = \gamma pqt \frac{d \log e}{d\sigma} - \sigma wt \frac{d \log e}{d\sigma} - \frac{d\alpha}{d\sigma} e^{\sigma} t - \alpha e^{\sigma} t \log e = 0 \qquad (32)$$

Total hours, t, enter proportionally because of the assumption of constant returns to scale. The last two terms sum to zero because hourly earnings with the initial level of effort per hour would not be affected by small changes in σ and α that kept workers on the same indifference curve.[56] Equation (32) then simplifies to

$$\frac{d \log e}{d\sigma} (\gamma pq - \sigma w) = 0 \qquad (33)$$

The first term gives the change in effort per hour from a compensated change in the elasticity; it is positive because a compensated change in σ necessarily changes e in the same direction. Therefore, since $pq = w$ by equation (31), equation (33) implies that

$$\sigma = \gamma;^{57} \text{ in particular } \sigma < 1 \text{ since } \gamma < 1, \qquad (34)$$

If there were no costs of monitoring the input of effort or time, firms would offer workers an earnings function that has the same elasticity with respect to effort as does the firm's output function. In effect, workers would be paid on a pure piece-rate basis; that is, they would receive the value of the output that they produce, for if $\sigma = \gamma$, a piece-rate equal to p has an implied earnings function that is identical to the explicit earnings function given by equation (29).[58]

A pure piece rate system is seldom used, even indirectly through the earnings function, presumably because it is costly to monitor the input of effort and time, or the quality of the product. Since time can usually be monitored much more cheaply than effort, errors in monitoring would be more serious when σ was larger, or when the earnings function was more effort-intensive. A given deviation between the effort supplied and the effort compensated would have a larger impact on hourly earnings when σ was large than when σ was small (it would have no impact when $\sigma = 0$). Therefore, firms would try to reduce monitoring errors when σ was large by spending more on monitoring the effort supplied by their employees.

Monitoring can be introduced formally into the analysis by specifying a monitoring cost function:

$$C = M(\sigma)t\zeta \quad , \tag{35}$$

where C is the total expenditure on monitoring, $M\zeta$ is the expenditure on monitoring the effort put into each hour, $\zeta \geq 0$ is a shift parameter, and the argument in the previous paragraph implies that the derivative of M with respect to σ is positive. The total income of a firm would be net of monitoring costs:

$$\pi = pQ - wt - C = p\beta e^{\gamma}t - \alpha e^{\sigma}t - M(\sigma)t\zeta \tag{36}$$

If π is differentiated with respect to t and σ, the optimality conditions are

$$pq = w + M\zeta \;, \tag{37}$$

and

$$\frac{d \log e}{d\sigma} (\gamma pq - \sigma w) = \frac{\partial M}{\partial \sigma} \zeta = M'\zeta \tag{38}$$

Equation (37) shows that hourly earnings are no longer equated to the marginal product of time, but are less by the expenditure on monitoring hourly effort. Workers can be said to pay for the cost of monitoring their effort because their earnings are sufficiently below their marginal product. Even if $\sigma = \gamma$, and firms were competitive, the implicit piece rate would not equal the price of the product, but would be less by the cost of monitoring each unit of effective time.[59]

Furthermore, equation (38) implies that the optimal σ would not equal γ, but would be less than γ whenever monitoring was costly, and would decline as the cost of monitoring increased.[60] An increase in monitoring costs increases the cost of using effort relative to time since the monitoring of time has been assumed to be costless. It is not surprising, therefore, that an increase in monitoring costs reduces σ: reduces the effort-intensity and raises thetime-intensity of the earnings function.

As a result of monitoring costs, the earnings function and a pure piece rate system differ in two ways: a unit of effective time is not paid the product price, and the effort elasticity of earnings does not equal the effort elasticity of output. Indeed, monitoring costs tend to[61] lower earnings both because a unit of effective time would be paid less than the product price, and because the earning elasticity would be less than the output elasticity. Consequently, the welfare of workers is reduced by monitoring costs because earnings are reduced below the compensation in a pure piece rate system.

Equation (38) shows that the optimal σ is determined not only by monitoring costs, but also by several other parameters. For example, an increase in γ raises σ:[62] an increase in the elasticity of output with respect to effort induces an increase in the elasticity of earnings because an increase in γ increases the marginal revenue from raising σ. Similarly, an increase in the size of the response of the supply of effort to a change in σ, or an increase in the supply of effort at given values of α and σ, also raise the optimal σ[63] because the marginal revenue from raising σ would be increased as long as $\gamma > \sigma$.

These results can be used to generalize the analysis to incorporate differences between workers and firms. As shown in Section III, the amount of effort supplied depends on different characteristics: for example, married men, single women, more energetic persons, richer persons, or those between the ages of say 25 and 50 supply more effort per hour than single men, married women, less energetic persons, poorer persons, or those younger and older, respectively. Even if firms were identical and used different types of workers in independent production processes having the same output and monitoring production functions, different types would not be offered the same earnings function. The second equation in the last footnote shows that workers supplying more effort per hour, such as married men or more energetic persons, would be offered a larger σ.[64]

Differences in the effort supplied per hour of work is one source of the inequality in hourly earnings referred to in the introduction to this paper. The analysis in Section III indicates that the inequality in effort would not be negligible in a labor force that differed substantially in marital status, age, sex, health, property income, or human capital. Moreover, the resulting inequality in hourly earnings would not simply mirror the inequality in effort. If all workers were offered the same earnings function, the inequality in earnings would be less than the inequality in effort because the elasticity of earnings with respect to effort is necessarily less than unity.[65] However, we have shown that workers supplying more effort are offered a larger elasticity, and probably also a larger compensation per unit of effective time. The inequality in hourly earnings is widened by the positive correlation between e, σ, and α; indeed, it could be widened sufficiently to greatly exceed the inequality in effort.[66]

Even if effort per hour were symmetrically distributed, earnings per hour would be skewed: negatively because the earnings elasticity is less than unity,[67] and positively because of the variation in this elasticity and the compensation per unit of effective time, and because of their positive correlation with effort.[68] Even with an earnings elasticity less than unity, the distribution of hourly earnings is likely to be more positively skewed than the distribution of hourly effort. Consequently, the positive skewness that is observed in actual distributions of earnings[69] can be explained without assuming that the distribution of effort is positively skewed.

It was shown earlier that an increased efficiency at monitoring or a larger elasticity of output with respect to effort would increase the earnings elasticity offered to workers. Consequently, if firms differed in their monitoring efficiency or output elasticities, the earnings functions offered even identical workers would be different in different firms. Since market equilibrium requires that identical workers will be equally well off in different firms, those offering larger earnings elasticities would also offer smaller compensations for each unit of effective time.[70] Although workers would be equally well off, they would not receive the same hourly earnings in different firms: earnings would be greater in firms offering larger elasticities because workers there have to be compensated for the additional effort per hour supplied. Consequently, differences among firms widens the observed inequality in hourly earnings by causing compensating differentials among identical workers.

When only workers or firms differ, either the same firm would offer different earnings functions to different types of workers, or different firms would offer compensating differentials to the same type. When both differ, however, the analysis is complicated by different sorting possibilities. would different types of workers be randomly distributed by type of firm, or Would workers supplying different amounts of effort be positively (or negatively) sorted with firms offering different earnings elasticities?

Our earlier analysis of the effect of an increase in effort on the optimal earnings elasticity suggests correctly that positive sorting is optimal because workers supplying more effort have a comparative advantage in firms offering larger elasticities: a larger elasticity is more attractive when the quantity of effort is larger.[71] Therefore, firms offering larger elasticities would be able to outbid other firms for workers supplying more effort, and firms offering smaller elasticities would be able to outbid other firms for workers supplying less effort.[72] Positive sorting of workers and firms increases the total output of an economy because workers and firms would be used more efficiently than in random (or negative) sorting. The effect of positive sorting on the distribution of hourly earnings is less clear: apparently it could either increase or decrease the inequality in earnings compared to a random sorting.[73]

V. The Production and Allocation of Effort by Slaves

The analysis of the demand for time and effort by firms applies to slave owners in their dealings with their slaves, except that a slave's supply function of time and effort differs from that of an employee because slaves cannot freely use their time and effort in household or other market activities. If there were no costs of monitoring the effort and time of slaves, and if investments cannot increase the stock of effort of slaves, the supply curves of time and effort would be completely inelastic with respect to compensation - at 24 hours per day and the given stock of effort - because slaves would be detected and presumably severely punished if they supplied less time or effort. An owner's income would be maximized by using all the time and effort of his slaves at productive activities, without compensating them in any way.[74]

To bring the analysis closer to that of firms hiring employees, assume that slave owners can invest in the effort (i.e., health, see Section IIIe) of their slaves by spending goods and the own time of slaves on the production of additional effort. A slave's supply of time and effort to productive activities would no longer be completely inelastic: the supply of time would be negatively, and the supply of effort would be positively related to the resources invested in producing additional effort. Since owners maximize their incomes, they would continue to invest in their slave's effort until the following conditions are satisfied:

$$p\ MP_E = p_x \frac{\partial x}{\partial E} \tag{39}$$

$$p\ MP_E = MP_t \frac{\partial t}{\partial E} \quad , \tag{40}$$

where the term on the left hand side gives the value to owners of an additional unit of effort.[75]

The right hand side of equation (39) gives the cost of the additional goods, and the right hand side of (40) gives the foregone value of the additional time, that are used to produce an additional unit of effort. In equilibrium, the value of the marginal product of effort must equal the marginal cost of producing effort with either goods or time. If the goods used to produce effort were "paid" to slaves as food, clothing, housing, medical care, etc., equation (39) would say that slaves are paid their marginal product in the sense that the value of additional effort would equal the marginal "wage" paid slaves in order to produce additional effort.

However, slaves are not paid their average product, even when there are no cooperating inputs, and this is why slaves are "exploited". By rewriting equation (39) and (40), the exploitation of slaves can be brought out more explicitly:

$$s_x = 1 - \frac{p_x x}{pQ} = 1 - \gamma\sigma_x \,, \tag{39'}$$

$$s_t = 1 - \frac{t_s}{t} = \frac{1 - \gamma}{1 - \gamma + \gamma\sigma_t} \,, \tag{40'}$$

where σ_x and σ_t are the elasticities of the production of effort with respect to goods and time respectively, γ is the elasticity of output with respect to effort, and t_s is the time slaves spend producing additional effort by sleeping, other rest, personal care, etc. The term s_x is between zero and unity, and measures the fraction of the output of slaves retained by owners; it has been called the "expropriation rate"[76] on goods. Similarly, s_t is also between zero and unity, and measures the fraction of the time of slaves that is used to produce output; it can be called the "expropriation rate" on time.

Since γ is less than unity, perhaps considerably less (see the discussion in Sections II and III), the expropriation rate on goods would be sizeable[77] unless σ_x, the elasticity of effort with respect to goods, was close to or above unity. Although no direct evidence is available, the elasticities of calorie, protein, or vitamin intake with respect to changes in incomes in poor countries are quite low, usually less than .25.[78] If these values are even a rough guide to the value of σ_x for slaves, then the expropriation rate on goods would be considerable; in particular, the rate for slaves in the American South would have been several times the estimate of 0.1[79] derived by Fogel and Engerman.[80]

If a slave owner knew the stock of effort of each of his slaves, the time and effort spent by each on all activities, and the elasticities of all outputs with respect to time and effort, he would determine a required input of time and effort for each slave and activity. To guarantee that these requirements or "quotas" were satisfied, he would threaten severe punishment to any slave who did not satisfy them. Of course, slave owners do not usually have such perfect knowledge because of the cost of monitoring the input of effort,[81] and of acquiring information about the effort capacities of different slaves.

Therefore, instead of closely tailoring quotas to each slave and activity, owners would combine a system of quotas with punishments and rewards. Given their accumulated knowledge of slave capacities, activities, and the influence of variables like weather, owners would establish quotas of effort per hour for each activity that would depend on the type of slave (prime male, child, pregnant woman, etc.), weather conditions, time of day, and so on. Slaves just satisfying these quotas would be neither rewarded nor punished.

Owners would recognize that the quotas might be too low for some slaves and activities because these slaves had larger than anticipated stocks of effort, or because productivity in these activities was greater than anticipated. In order to encourage slaves to exceed their quotas when that is possible instead of hoarding any extra effort for personal use, owners would give extra food, clothing, cash, time off, etc. - that is, "incentive pay" - to slaves that exceeded their quotas.

Owners would discourage slaves from not meeting their quotas by punishing them when they are discovered. Slaves cannot be punished by withdrawing their incentive pay because there is no incentive pay when the quotas are just met. Consequently, they have to be punished in other ways; for example, by whippings and other torture, imprisonment, or starvation.[82] Since owners know that their quotas would be too high under certain conditions, punishments would often be mild - extenuating circumstances could absolve a slave of much guilt -, and would become more severe when transgressions were more serious.

Stated more formally, the optimal system of punishments and rewards has the following form when there are significant costs of monitoring and evaluating the performance of slaves:

$$
\begin{aligned}
I &= \alpha_I (e - e_0)^{\sigma_I} \text{ when } e \geq e_0 \\
P &= \alpha_p (e_0 - e)^{\sigma_p} \text{ when } e < e_0 ,
\end{aligned}
\tag{41}
$$

where e_0, the required effort per hour, depends on the type of slave and activity, I is the expected incentive pay when effort exceeds e_0, P is the expected punishment when effort is less than e_0, σ_I and σ_p are the elasticities of the pay and punishment segments, respectively, and α_I and α_p are compensation and punishment respectively, for each unit of $(e - e_0)^{\sigma_I}$ or $(e_0 - e)^{\sigma_p}$. The elasticities that maximize an owner's income would be larger, the more elastic is output with respect to effort, and the more elastic are the supply curve of slaves with respect to their incentive pay and punishment.[83]

Both incentive pay and actual (as opposed to the threat of) punishment would be unnecessary if slave owners had complete knowledge of each slave and activity, and are introduced to protect owners against some of the adverse consequences of incomplete knowledge. In particular, they protect against deviations between actual and optimal "quotas" on effort per hour: incentive pay protects against quotas that are too low because they encourage slaves to surpass their quotas, and a flexible system of punishments protects against quotas that are too high, because they separate malingering from owner error. Consequently, incentive pay and punishment should both be more common when owners have less knowledge of their slaves capacities and the conditions of production in different activities.

This conclusion is relevant to the controversy over the incidence of incentive pay and physical punishment of slaves in the South. One side (see Fogel-Engerman) claims that incentive pay was common and physical punishment such as whippings, not common, whereas the other side (see Stampp and Gutman-Sutch in Stampp) deemphasizes incentive pay and claims that physical punishment was common. If our analysis is correct, both sides are right and wrong:

incentive pay would be common when punishment was common, and rare when punishment was rare.

Not only would incentive pay and physical punishment be eliminated when owners completely knew each slave and activity, but so too would decision-making by slaves since the allocation of all their time, effort, and goods would be completely determined by their owners. Slaves have some freedom to choose how to allocate their resources when owners have limited knowledge, and offer slaves the pay-punishment given by (41). The utility maximizing analysis developed in Sections II and III for free employees would then be applicable to slaves, except that employees are not physically punished,[84] and are compensated for all their time and effort; they are not given quotas to be met without compensation. Still, many of the results on the effects of changes in productivity, the stock of effort, and earnings elasticities on the supply of effort and time by employees apply to slaves as well.

FOOTNOTES

[1]Since

$$Q = \frac{\partial Q}{\partial E} E + \frac{\partial Q}{\partial K} K + \frac{\partial Q}{\partial t} t \ ,$$

$$\varepsilon_{QE} + \varepsilon_{QK} + \varepsilon_{Qt} = 1,$$

where ε_{QE} is the elasticity of Q with respect to E, etc. Clearly,

$$\varepsilon_{QE}, \varepsilon_{QK}, \varepsilon_{Qt} < 1 \ .$$

[2]If effort and other inputs are "complements" - have positive cross-derivatives -, diminishing marginal product follows because

$$\frac{\partial^2 Q}{\partial E^2} E + \frac{\partial^2 Q}{\partial K \partial E} K + \frac{\partial^2 Q}{\partial t \partial E} t = 0$$

Looked at differently, if the elasticity of output with respect to effort is either constant or declining as effort increases, then diminishing marginal product is implied because this elasticity is less than unity. Since

$$\frac{\partial Q}{\partial E} = \gamma_E \frac{Q}{E} \quad ,$$

where $\gamma_E < 1$ is the elasticity, then

$$\frac{\partial^2 Q}{\partial E^2} = \frac{\gamma_E Q}{\partial E^2} (\gamma_E - 1) + \frac{\partial \gamma_E}{\partial E} \frac{Q}{E} < 0 \quad \text{if } \frac{E}{\partial E} < 0$$

[3]With separate prices for effort and time, the earnings of any employee would be

$$I = w_e E + w_t t,$$

where w_e and w_t are the unit prices of effort and time respectively, and E and t are the amounts of each that he supplies. Then even if $t = 0$, $I = w_e E > 0$ if $E > 0$, and even if $E = 0$, $I > 0$ if $t > 0$. Presumably, however, I should equal zero when <u>either</u> E or $t = 0$, regardless of the value of the other input.

[4] Stated more technically, the earnings function is the "envelope" of the offer functions of different firms. Even though a firm offers less than the maximum earnings at a particular level of effort, it can survive if it can offer the maximum at other levels, and can employ persons at these levels. A further discussion of this "sorting" of firms and workers can be found in Section IV.

[5] A formally similar "interaction" between the quantity and quality of children has been treated by Becker and Lewis (1973) and Becker and Tomes (1976).

[6] For the present only single person households are considered; later I treat some consequences of the division of labor in multi-person households.

[7] Indeed, sleep is more effort-producing than effort-using (see the discussion in Section IIIc).

[8] These estimates and their interpretation are discussed more systematically in the next section.

[9] Beesley's estimates (1965) rise from about 30 percent of hourly earnings for lower income persons to 50 percent for higher income persons; similar results were obtained by Lisco (1967). According to Becker (1965), the time spent in commuting is valued at about 40 percent of hourly earnings. Gronau (1970) separately estimates the time cost of personal and business airplane trips; he concludes that business time is valued at about hourly wage rates whereas personal time is considered free. Also see McFadden.

[10]Not only have they been difficult to explain, but they also have major implications for benefit-cost evaluations of super-highways, supersonic airline travel, and other fast modes of travel because the major benefit of these modes is the saving in time costs.

[11]Whether or not traveling per se provides positive or negative utility is intrinsically a separate issue from the shadow cost of travel time, just as the utility from eating is separate from the cost of the time used in eating. However, the methods that have been used to estimate the cost of travel time are affected by whether or not travelling provides utility.

[12]If

$$w_m \frac{(1-\sigma_m)}{(1-\sigma_i)} = kw_m ,$$

then

$$\sigma_m = (1-k) + k\sigma_i \geq 1-k.$$

If $k = .5$, $\sigma_m \geq .5$; if $k = .4$, $\sigma_m \geq .6$.

[13]Recall that

$$I_m = \alpha_m e_m^{\sigma_m} t_m = \alpha_m E_m^{\sigma_m} t_m^{1-\sigma_m} ,$$

where I_m is earnings. If $\sigma_m \geq 1/2$, then

$$\frac{\partial I_m}{\partial E_m} \cdot \frac{E_m}{I_m} = \sigma_m \geq \frac{\partial I_m}{\partial t_m} \cdot \frac{t_m}{I_m} + 1 - \sigma_m .$$

[14]If the value of time equalled 40 percent of the wage rate,

$$.6 \leq \sigma_m \leq .8 \text{ implies } 0 \leq \sigma_i \leq .5$$

[18]The pioneering studies are by Gronau and Heckman.

[19]Since hours per week often exceeded 53 prior to the 1920's in the United States, Denison presumably is assuming that both workers and firms either were stupid or unaware of the negative effects beyond 53. Since both assumptions are unreasonable, he probably overstated the negative effect of hours beyond 53, and between 43 and 53 hours as well.

[20]This assumption is made by Freudenberger and Cummins.

[21]If there is only a single household activity, then

$$e_m t_m + e_h t_h = E,$$

and

$$\frac{\partial e_m}{\partial t_m} = - \frac{(e_m - e_h)}{t_m}$$

By equation (22), $e_h < e_m$ if $\sigma_m > \sigma_h$. Hence

$$\frac{\partial e_m}{\partial t_m} < 0.$$

[22]Indeed, the effort spent on each household and working hour would change by the same percentage. However, the effort spent on the <u>average</u> household hour would probably change by a still larger percentage because time is mainly added to or subtracted from household activities that are relatively effort-intensive (see Section IIa).

[15]Consider two modes A and B that take the same time, t, but have different direct costs, x_a and x_b, and effort elasticities σ_a and σ_b. Assume that the more effort-intensive activity A has lower direct costs (otherwise it would never be chosen), and that initially A is chosen; that is, the sum of direct and indirect costs would be lower for A:

$$x_a + wt \frac{(1 - \sigma_m)}{1 - \sigma_a} < x_b + wt \frac{(1 - \sigma_m)}{1 - \sigma_b} ,$$

or

$$x_b - x_a > \frac{(1 - \sigma_m)\ (\sigma_a - \sigma_b)\ wt}{(1 - \sigma_a)\ (1 - \sigma_b)}$$

The less effort-intensive mode B would be chosen if hourly earnings, w, increased sufficiently to reverse this inequality.

[16]Beesley ( ) and Lisco ( ) claim to find, however, that persons with higher earnings value their commuting time at a larger fraction of their hourly earnings. I say "claim" because their estimates are based on the assumption that hours worked are the same for persons with different earnings, although in fact hours worked and earnings are significantly positively related (see the evidence for the United States in Section IIIc). Their estimates are biased, therefore, toward a positive relation between earnings and the relative value placed on commuting time. McFadden does not find strong evidence that persons with higher earnings value their commuting time at a larger fraction of their earnings.

[17]See equation (16).

[23]From footnote 21,

$$\frac{\partial e_m/e_m}{\partial t_m/t_m} = -(1 - \frac{e_h}{e_m}) > -1$$

[24]Since

$$\frac{dw_m}{w_m} = \sigma_m \frac{de_m}{e_m} ,$$

$$\frac{dw_m}{w_m} / \frac{dt_m}{t_n} = \varepsilon_m = -\sigma_m (1 - \frac{e_h}{e_m}) >> -1.$$

If $\sigma_m = 3/4$ and $\sigma_h = 1/2$ because the shadow price of household time equals 1/2 of hourly earnings, then

$$1 - \frac{e_h}{e_m} = 1 - \frac{\sigma_h(1 - \sigma_m)}{\sigma_m(1 - \sigma_h)} = 2/3 ,$$

and $$\varepsilon_m = 3/4 \cdot 2/3 = 1/2.$$

[25]See the evidence accumulated by Paul Douglas (Chapters XI-XII).

[26]See Mincer, Rosen, and Wigner.

[27]If the stock of effort increased by 10 percent, and if hours worked were unchanged, effort per hour would also increase by 10 percent. Since hours worked would tend to increase (see the analysis in the next section), effort per hour must increase by less than 10 percent and could decrease. However, since the elasticity of effort per hour with respect to hours worked is about -1/2 (see footnote 24), effort per hour would increase as long as the elasticity of working hours with respect to the stock of effort was less than +2, which is likely to be true.

[28]The reasons for this difference are discussed in Section IV.

[29]That is,

$$\alpha_{m_2} = \alpha_{m_1} \hat{e}_{m_1}^{\sigma_{m_1} - \sigma_{m_2}} ,$$

where $\hat{e}_{m_1}$ is the optimal level of e_m in job 1. The proof is identical to the proof that a constant Laspeyres price index would keep a consumer on the same indifference curve for small changes around an initial equilibrium position. If hourly earnings were the same when $e_{m_2} = \hat{e}_{m_1}$, a worker could obtain the same goods, household effort, and hence utility in job 2 as he does in job 1 by supplying $\hat{e}_{m_1}$ units of effort per hour and $\hat{t}_{m_1}$ hours of work to job 2. Of course, he would react to the differences in σ_m and α_m by supplying different quantities of e_m and t_m to job 2, but he would still receive the same utility if the differences in σ_m were small.

[30]By the last two conditions of equation (14),

$$U_{E_i} = \tau w'_m = \tau \sigma_m w_m e_m^{-1} ,$$

where U_{E_i} is the marginal utility of the total effort spent on the ith household activity. An increase in σ_m, with w_m and e_m held constant, would increase w'_m , the marginal cost of the effort used in activity i.

[31]Recall from Section IIIb that the marginal cost of using time at the ith household activity is

$$\hat{w}_i = w_m \frac{(1-\sigma_m)}{1-\sigma_i}$$

[32]This implication is consistent with Marshall's conclusion that "As a general, though not universal rule,...work is more intense when paid by piece, than when paid by time; and, insofar as this is the case, short hours are specially suited to industries in which piece-work prevails." (p 693)

[33]The inequality in energy is dramatically conveyed in the following preface to a biography of Gladstone: "Lord Kilbracken, who was once his principal private secretary, said that if a figure of 100 could represent the energy of an ordinary man, and 200 that of an exceptional man, Gladstone's energy would represent a figure of at least 1,000." (See P. Magnus, Gladstone (John Murray, 1954) p. xi.) I owe this reference to George Stigler.

[34]The elasticity of working hours with respect to an increase in the stock of effort can be expressed as

$$n_{t_m E} = \frac{t_m}{t_m R} \left[(\sigma_m - \sigma_h)\, x\sigma_c - \sigma_m\, (p_x x - v)\, \mu_t + \sigma_h\, p_x x\, \mu_x \right] ,$$

where t_h and x are the total time and goods used in the household, μ_t and μ_x are the full income elasticities of t_h' and x respectively, σ_c is the elasticity of substitution between x and t_h' in the derived utility function, and R is a positive function of various parameters. The substitution effect is essentially given by

$$(\sigma_m - \sigma_h)\, x\, \sigma_c > 0 .$$

The income effect is given by

$$\sigma_h\, p_x\, x\, \mu_x - \sigma_m\, (p_x\, x - v)\mu_t \gtreqless 0 .$$

It is less than zero only if

$$\frac{\sigma_h}{\sigma_m} < k_e \frac{\mu_t}{\mu_x} ,$$

where k_e is the share of earnings in money income. The discussion in this footnote is based on notes of H. Gregg Lewis.

35 Since

$$\log I = \log \alpha_m + \sigma_m \log e_m (E, t_m) + \log t_m,$$

then if α_m and σ_m are not affected,

$$\frac{d \log I}{d \log E} = \sigma_m \frac{\partial \log e_m}{\partial \log E} + \sigma_m \frac{\partial \log e_m}{\partial \log t_m} \frac{d \log t_m}{d \log E} + \frac{d \log t_m}{d \log E}$$

$$= \sigma_m + \frac{d \log t_m}{d \log E} ((1 - \sigma_m (1 - \frac{e_h}{e_m}))$$

(frcm footnote 23). If $\sigma_m = 3/4$, and $\sigma_h = 1/2$, then $1 - \frac{e_h}{e_m} = 2/3$, and I would increase by a larger percentage than E if $\frac{d \log t_m}{d \log E} > 1/2$; if σ_m also increased when E increased, the elasticity required of t_m would be even smaller.

36 Since the coefficient of proportionality need not be the same for different persons, differences in their health cannot necessarily be inferred from differences in their stocks of effort.

37 See, for example, Grossman

38 Michael Grossman, in his penetrating analysis of health, assumes that ill health raises sick time; that is, reduces the time available for work or for other household activities. This formulation assumes, it does not explain, why sick time is spent at home rather than at work. Grossman's formulation readily explains why ill-health reduces hours worked but not why it also reduces hourly earnings. However, if persons in ill-health invest less in other kinds of human capital (see the argument in Grossman 1975, and my discussion later on), they would have lower hourly earnings after investments are completed (and higher earnings when investments are beginning).

[39] In Grossman's formulation (op. cit.), on the other hand, health is neutral between the market and household sectors.

[40] See the evidence on calories cited earlier.

[41] I am assuming that certain units of different input are devoted <u>exclusively</u> to the production of additional effort. The analysis can, however, also be readily developed for "joint production"; when inputs, like a good diet, directly produce both effort and household commodities.

[42] If σ_m increased, w'_m is directly increased; if σ_i increased, w'_m is indirectly increased through the decrease in e_m.

[43] Grossman (1975) presents evidence both that healthier persons invest more in schooling, and that schooling encourages investment in health.

[44] See Auster-Levison , Grossman , and Fuchs .

[45] Earnings may also increase because the effort devoted to each working hour increases, perhaps as a result of a decrease in household effort elasticities (see the discussion in Section IIId). An increase in effort per hour reduces the incentive to invest in health because its marginal product and hence its shadow price is reduced. This negative relation between health and effort per hour has been recognized in the health literature, and called "ruining one's health" by overwork, or "burning oneself out" (see the following discussion of life cycle variations in effort).

[46]The shadow price of effort to a person not in the labor force equals

$$\frac{\varepsilon}{\tau} = p_x \frac{U_{t_h'}}{U_x} w_h' ,$$

where the second term on the right is the marginal rate of substitution in his utility function between effective time and goods, and w_h' is the marginal productivity of additional effort in household activities. An increase in property income raises the shadow price of effort because it raises the marginal rate of substitution: time and effort become scarcer relative to goods.

[47]The optimal path has been frequently derived for other kinds of human capital (see, for example, Ben-Porath, Ghez-Becker). These derivations depend on the finiteness of life, but if health is related to the stock of effort, the length of life cannot be taken as given, but would be partly dependent on decisions about the optimal stock of effort (see Grossman). One could say that the life of a person would end when his stock of effort fell below a prescribed level. His stock of effort (and health) would eventually begin to decline because either the rate of depreciation or the marginal cost of producing effort eventually would begin to rise with age.

[48]Each single household allocates its resources the same way because they are, by assumption, identical households. Each mate would continue this allocation when married because this allocation would continue to satisfy all the optimizing conditions of equation (14). Since the household production functions have constant returns, the output of each commodity in the married household would be twice the output in each single household.

[49]The efficient combination can dominate (in the sense discussed by Becker [1973]) any other sorting.

[50]In 1969, married men with spouse present who were over 35 years old earned 28 percent more per hour than never married, and 17 percent more than separated, divorced, or widowed men over age 35 (see Fuchs, 1974, Table 1). The annual earnings in 1966 of married men with spouse present were 17 percent higher than that of separated and divorced persons the same age, having the same years of schooling, and the same number of weeks worked (unpublished calculations from Becker-Landes-Michael).

[51]See

[52]Even if marriage has a much greater effect on the health of men, the health of married women could be better than that of married men because single women would have better health than single men since single women are more skilled (household) producers of health.

[53]Other inputs, such as physical capital, are ignored; see the discussion in Goldberg

[54]If an increase in α reduced the supply of effective labor time - if the supply curve of effective time were "backward bending" -, a rise rather than a fall in α and p would be equilibrating only if the negatively sloped demand curve was flatter than the negatively sloped supply curve.

[55]Since firms are assumed to be indifferent to the distribution of hours worked among identical workers supplying the same effort per hour, it is not necessary to consider the effect of changes in α and σ on the hours supplied by each worker.

[56]This follows from a basic theorem on price indexes (see footnote 29). If hourly earnings did not change,

$$\frac{dw}{d\sigma} = 0 = \alpha e^{\sigma} \log e + \frac{e^{\sigma} d\alpha}{d\sigma},$$

or

$$\frac{d\alpha}{d\sigma} = - \alpha \log e.$$

[57]This result also follows when each firm is a monopoly in the product market. Then $p\bar{q}$ would be replaced by $pq(1 + \frac{1}{\varepsilon_D})$ in equations (31) and (33), where ε_D is the elasticity of the demand curve for the firm's product. The modified equation (33) would still imply that $\sigma = \gamma$.

This result would not necessarily follow, however, if physical capital or other inputs affected output and costs. The μ effect on the relation between γ and σ depends on how they enter the cost and output functions (see Goldberg).

[58]That is, if $\sigma = \gamma$,

$$pQ = p\beta e^{\gamma} t = p\beta e^{\sigma} t = \alpha e^{\sigma} t$$

since equation (31) implies that $\alpha = p\beta$ when $\sigma = \gamma$. If firms were monopolists the piece rate would not equal p, but $p(1 + \frac{1}{\varepsilon_D}) < p$, and

$$\alpha = p \left(1 + \frac{1}{\varepsilon_D}\right) \beta < p\beta;$$

otherwise, the interpretation of $\sigma = \gamma$ is the same for monopolists.

[59]If $\sigma = \gamma$, equation (37) implies that

$$\alpha = p\beta - \frac{c}{e^{\sigma}t},$$

where $e^{\sigma}t$ is the amount of effective time.

[60]By differentiating (38) with respect to the monitoring shift parameter ζ, one obtains

$$\frac{d\sigma}{d\zeta} = \frac{M'}{\Delta},$$

where

$$\Delta = \frac{1}{t}\frac{\partial^2\pi}{\partial\sigma^2} = \frac{d^2 \log e}{d\sigma^2}(\gamma pq - \sigma w) + \left[\frac{d \log e}{d\sigma}\right]^2(\gamma^2 pq - \sigma^2 w) - \frac{d \log e}{d\sigma} w - M''\zeta < 0$$

by the second order conditions for a maximum position. Hence

$$\frac{d\sigma}{d\zeta} < 0 \text{ because } M' > 0,$$

or an increase in monitoring costs reduces σ. Since $\sigma = \gamma$ when $\zeta = 0$ because equations (37) and (38) then reduce to (32) and (33), σ must be less than γ for all $\zeta > 0$, and must get smaller as ζ gets larger.

[61]I only say "tend to" because the effect on earnings is also determined by the elasticity of the aggregate supply of effective time.

[62]If γ depends positively on a parameter s, then

$$\frac{\partial\sigma}{\partial s} = -\frac{\frac{d \log e}{d\sigma}(pq + pq \log e)\frac{d\gamma}{ds}}{\Delta} > 0$$

since $\Delta < 0$ (see footnote 60) and $\frac{d\gamma}{ds} > 0$.

[63]

$$\frac{\partial\sigma}{\partial\mu} = - \frac{\frac{d}{dv}\left(\frac{d \log e}{d\sigma}\right)(\gamma pq - \sigma^2 w)}{\Delta} > 0$$

since $\Delta < 0$ and both terms in the numerator are positive. Similarly,

$$\frac{\partial\sigma}{\partial v} = - \frac{\frac{d \log e}{dv}(\gamma^2 pq - \sigma^2 w)}{\Delta} < 0$$

since $\gamma > \sigma$, and $\frac{d \log e}{dv} > 0$. I am assuming that an increase in v increases e but does not change $\frac{d \log e}{d\sigma}$.

[64]The equation in footnote 59 suggests that an increase in effort per hour raises α if it lowers monitoring costs per unit of effective time. The direct effect of an increase in effort lowers these costs, but the indirect effect raises them through the induced increase in σ. Although the net effect on α is not completely clear, an increase in effort probably also tends to increase α.

[65]If earnings equalled

$$I = ae_i^{\sigma},$$

where a and σ were constants, then

$$V(\log I) = \sigma V(\log e) < V(\log e) \text{ since } \sigma < 1,$$

where V refers to the standard deviation of the particular distribution.

[66]If α, σ, and e were uncorrelated with each other, the variance in the log of earnings per hour would be

$$V^2 \log(I) = V^2(\log \alpha) + \bar{\sigma}^2 V^2(\log e) + (\overline{\log e})^2 V^2(\sigma) + V(\sigma)V(\log e),$$

where $\bar{\sigma}$ and $\overline{\log e}$ are the means of σ and log e, respectively. If the variation in log α and σ were not negligible, the variance in log I could easily exceed $\bar{\sigma}^2 V^2(\log e)$, the variance in the log of earnings when α and σ are the same for everyone. Moreover, we have shown that σ and e (and probably α too) would not be uncorrelated, but strongly positively correlated, and this raises the variance in log I much further, again if the variation in log α and σ were not negligible (see Goodman). Furthermore, random variation in α and e, and even more the positive correlation between σ and e, would induce a response in e that would widen the inequality in e. (The positive correlation between α and $\dot{e}$ might, however, somewhat reduce the variation in e.)

[67]If α and σ are positive constants, with $\sigma < 1$, the distribution of $I = \alpha e^{\sigma}$ would be less skewed to the right than the distribution of e; in particular, if e were symmetrically distributed, I would be skewed to the left.

[68]The distribution of $I = \alpha e^{\sigma}$ would be positively skewed if α and e^{σ} were symmetrically distributed and uncorrelated with each other; the positive skew would increase as the positive correlation between α and e^{σ} increased (for a proof with normal distributions, see Craig). If e were symmetric, variation in σ, and especially a strong positive correlation between e and σ would increase the positive skewness in e^{σ}, and thereby in I.

[69]The ratio of the mean to median weekly earnings, one measure of skewness, was 1.13 for ever married white males aged 15-65 with positive earnings in 1966; this ratio rose from 1.06 for males 15-24 to 1.17 for males 55-64. Two other measures of skewness - the ratio of the third moment around the mean to the cube of the standard deviation, and the difference between the deviation between the third quartile and the median and the deviation between the median and the first quartile divided by the sum of these deviations - also show moderate positive skewness (all measures calculated from data prepared for Becker-Landes-Michael).

[70] The relation between α and σ is given by the following differential equation:

$$\frac{d\alpha}{d\sigma} = -\alpha \log e,$$

where e is the optimal supply of effort at the point where the derivative is evaluated (see fn 56).

[71] Consider two types of firms with output functions $\beta_1 e^{\gamma_1}$ and $\beta_2 e^{\gamma_2}$, and product prices p_1 and p_2 respectively. If, to simplify, monitoring costs are assumed to be zero in both types, then $\alpha_i = p_i\beta_i$, and $\sigma_i = \gamma_i$, $i = 1, 2$ (see equation (34) and footnote 58), and the earnings function offered all workers would be $I_1 = p_1\beta_1 e^{\sigma_1}$ in type 1 firms, and $I_2 = p_2\beta_2 e^{\sigma_2}$ in type 2 firms. If there are two types of workers who supply e_1 and e_2 units of effort per hour when employed in firm type 1, and $e_1(1 + s_1)$ and $e_2(1 + s_2)$ units when employed in firm type 2, the ratio of the hourly earnings of each type of worker in the two types of firms is

$$r_1 = \frac{p_2\beta_2 e_1^{\gamma_2}(1 + s_1)^{\gamma_2}}{p_1\beta_1 e^{\gamma_1}} = \frac{p_2\beta_2}{p_1\beta_2} e_1^{\gamma_2-\gamma_1} (1 + s_1)^{\gamma_2}$$

$$r_2 = \frac{p_2\beta_2 e_2^{\gamma_2}(1 + s_1)^{\gamma_2}}{p_1\beta_1 e_2^{\gamma_1}} = \frac{p_2\beta_2}{p_1\beta_1} e_2^{\gamma_2-\gamma_1} (1 + s_2)^{\gamma_2}$$

If the elasticity of e with respect to σ and α were the same for both types of workers - if $s_1 = s_2$ -, then $r_2 > r_1$ if $e_2 > e_1$ and $\gamma_2 > \gamma_1$: workers supplying the larger quantity of effort have a comparative advantage in firms having the larger output elasticity. These equations indicate further that comparative advantage depends also on the elasticity of supply: if e_1 and e_2 were equal but $s_2 > s_1$, type 2 workers would have a comparative advantage in type 2 firms. The same conclusions tend to hold when there are positive monitoring costs.

[72]If the supply of effort were not responsive to the terms of remuneration, a worker would simply choose the firm that pays him the highest hourly earnings. Then some workers would be employed in each type of firm considered in the previous footnote only if

$$r_2 \geq 1 \text{ and } r_1 \leq 1,$$

or

$$e_2^{\gamma_1 - \gamma_2} \leq \frac{p_2\beta_2}{p_1\beta_1} = \frac{\alpha_2}{\alpha_1} \leq e_1^{\gamma_1 - \gamma_2} \quad .$$

Since r_2 is strictly greater than r_1, both could not equal 1; hence at least one type of worker would be employed exclusively in one type of firm. The same kind of conclusion holds when there are many types of workers and firms: most workers (and firms) would be specialized to a small number of types of firms (or workers). Moreover, the supply of effort can be permitted to respond to changes in the terms of remuneration without altering this conclusion about specialization, as long as the elasticity of response (s_1 and s_2 in the previous footnote) did not greatly differ among types of workers.

[73]The ratio of the hourly earnings of "rich" to "poor" workers in the example considered in the previous two footnotes is bounded by

$$\left(\frac{e_2}{e_1}\right)^{\gamma_1} \leq \frac{I_2}{I_1} \leq \left(\frac{e_2}{e_1}\right)^{\gamma_2} \qquad \text{if } s_1 = s_2$$

because of the bounds on $\frac{\alpha_2}{\alpha_1}$ derived in footnote 72. A random sorting would tend to have larger inequality if $\frac{I_2}{I_1}$ were close to its lower bound, and smaller inequality if $\frac{I_2}{I_1}$ were close to its upper bound.

[74] An owner maximizes

$$I = p\beta E^{\gamma} t^{1-\gamma}$$

subject to $E \leq E_0$ and $t \leq t_0$. The marginal products of effort and time are

$$\frac{\partial I}{\partial E} = \gamma I E^{-1}, \text{ and } \frac{\partial I}{\partial t} = (1 - \gamma) I t^{-1}$$

Since $0 < \gamma < 1$ (see Section II), both marginal products are strictly positive and hence all the available time and effort, t_0 and E_0, would be used.

[75] An owner mazimizes

$$I = p\beta E^{\gamma} t_m^{1-\gamma} - p_x x$$

subject to

$$E = E(x, t_s),$$

and

$$t = t_m + t_s ,$$

where t_s is the own time of slaves, x the goods and services used to produce additional effort, and t_m the time of slaves devoted to the production of output. Maximizing I with respect to x and t_s yields equations (39) and (40).

[76] See Fogel and Engerman

[77] It must be positive since $\gamma\sigma_x < 1$ (if γ and σ_x are constants) by the second order conditions that are associated with the first order income-maximizing conditions given by (39') and (40').

[78] See Selowsky

[79]The value of σ_x for slaves might be much higher than these income elasticities in poor countries because owners have considerable control over how slaves spend their "incomes", and would provide nutritional food, and other productive goods (see the discussion of the slave's diet in Fogel-Engerman, and Sutch). By the same token, however, the utility slaves get from a given "income" would be less than that of free persons because slaves would be forced to trade off utility for productivity. Consequently, the goods consumed by slaves would be an overestimate of their real income or well-being. In addition, since slaves are forced to work longer hours than they want to, the fraction of time spent producing output, t_s, is suggested as a second measure of exploitation. This measure also understates the exploitation of slaves because owners direct the "free" time of slaves to effort-creating activities, not to the utility-creating activities desired by slaves.

[80]Their estimate is not based on the income maximizing conditions given by (39) and (40), but on slave prices and expenditures on slaves. Although David-Temin severely criticize the method and estimate of Fogel-Engerman, they too do not use the income maximizing conditions.

[81]I assume that owners know the allocation of time of their slaves (a similar assumption is applied to firms in Section III).

[82]I have argued (Becker, 1968) that physical punishment of free persons committing crimes is often unwise because the same deterrence can be achieved by fines or other monetary transfers, and they are cheaper to society. Slaves not meeting their quotas could not be punished by monetary transfers alone because they would not receive any incentive pay. If owners were spending goods and "free" time on slaves to increase their health (see the discussion

of equations (39) and (40)), they could punish slaves by reducing these expenditures. Owners would also be punished, however, because they are harmed by any reduction in their slaves' health; for example, starvation is a punishment to slaves but it also harms owners because their slaves would be less energetic. Hence starvation and other reduced expenditures on health do not simply transfer resources to owners, like fines do, but also harm owners, like other physical punishments do.

[83] Section III contains proofs of these and other propositions for firms hiring employees; similar proofs can be developed for slave owners.

[84] Essentially, because physical punishment is not efficient (see fn 82).