DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES

CALIFORNIA INSTITUTE OF TECHNOLOGY

PASADENA, CALIFORNIA 91125

INCENTIVES AND THE OPTIMALLY IMPERFECT INFORMATION STRUCTURE

Ross M. Miller

SOCIAL SCIENCE WORKING PAPER 383

May 1981

Revised July 1981

ABSTRACT

This paper constructs the optimal incentive structure for an economy through the simultaneous choice of the optimal income tax and the optimal information structure. An information structure in the economy is given by the proportion of workers whose productivity is directly observable by employers relative to the number who can only convey productivity through a signal. A second-best situation exists because the government has distributional goals and the tax it levies to achieve these goals depends on income and not on ability.

The main result of this paper is that the optimal information structure is the one where the productivity of no worker is directly observable by employers, i. e., all workers must signal productivity. This result comes about because the income tax may be adjusted when productivity is not observable to achieve any distribution of incomes that is possible when produotivity is observable. In addition, the incentive effects of signaling serve to offset the disincentive effects of the income tax.

INCENTIVES AND THE OPTIMALLY IMPERFECT INFORMATION STRUCTURE

Ross M. Miller

I. INTRODUCTION

This paper constructs the optimal incentive structure for an economy by considering not only the optimal payoff (tax) schedule for the economy but also by determining the optimal information structure for the economy. The vast literature in optimal taxation theory that has been done along the lines of Mirrlees' (1971) seminal work has generally assumed that the information structure of the economy is given and only the tax schedule is subject to adjustment by the taxing authority.

This paper derives the optimal income tax and information structure for an economy where employees choose a productivity-enhancing activity (referred to as education) where that activity may also serve as a signal of productivity.[1] The productivity of at least some of the employees (including the productivity due to education) is capable of being observed directly and noiselessly by employers, but employers can only infer the productivity of the other employees from their education. An information structure for the economy may then be defined in terms of the number of workers whose productivity is directly observable by employers relative to the number who can only signal productivity.

The main result of this paper is that in an economy with two productivity types, it is optimal to allow no employer to use information about productivity, i.e., it is optimal to make all workers go through the signaling process.

It has long been known that restrictions on the transmission of information may be socially beneficial. Hirshleifer (1971), Arrow (1973) and Spence (1974b) show that that this will be the case when the information generating aotivity is not productive in the usual sense and serves only to redistribute income in a way that increases the spread of incomes. In the context of the principal and agent relationship Green (1979) and Green and Stokey (1980) show that a more informative information structure may leave both principal and agent worse off. In general, however, more information can benefit both principal and agent, even if the information is extremely noisy, as in Holmstrom (1979) and Shavell (1979).

The important feature which differentiates this paper from previous work is that the information structure is chosen in a setting where an income tax may be optimally chosen. The fact that the choice of information structure has positive distributional effects stands in striking contrast to the work of Hylland and Zeckhauser (1979) and Shavell (1981), where distributional objectives are satisfied through the choice of the income tax and cannot be achieved otherwise. Hylland and Zeckhauser show that government programs providing benefits that depend on an individual's pre-tax (or post-tax) income should be designed to maximize net benefits (i.e., they should not be used to redistribute income). Their result does not extend to the case where benefits depend on ability; hence, it does not apply to the choice of information structure as presented in this paper. A more informative information structure in this paper will benefit high ability (productivity) workers at the expense of low ability (productivity) workers independent of their incomes.

This paper starts with the basic problem of optimal income taxation theory: the achievement of distributional goals in an economy causes a change in individual incentives that results in inefficiency. This inefficiency is a result of the standard assumption in the theory of optimal income taxation that ability cannot be taxed. In this model the inefficiency manifests itself as an underinvestment in education. Inducing the use of education as a signal by making productivity unobservable increases the incentives to purchase education and counteracts the disincentive effect of the income tax. Alternatively, as the proof of the main result demonstrates, a (nonlinear) income tax can be structured not only to remedy the inefficiency of signaling behavior, as Spence (1974a) shows, but also can be adjusted to achieve any distribution goal that is achievable when productivity is observable. Since denying the use of information about productivity may viewed as the revocation of a property right of the more productive workers, income is more easily redistributed when these workers are forced to signal productivity. The government may also induce signaling behavior without affecting the observability of productivity by requiring that employers make wage decisions that are based solely upon the signal, as is the case with affirmative action rules.

This paper begins by comparing the case the productivity of all workers is observable with the case where all workers must signal. The extension to intermediate information structures is facilitated by

examining the extreme cases first. Section II develops the productivity observable case and Section III develops the education observable (signaling) case. In Section IV it is proved that the signaling information structure always yields as much welfare as the observable productivity information structure. Section V gives a numerical example which illustrates the result in Section IV. Section VI states and proves the general theorem that the signaling information structure yields at least as much welfare as the intermediate information structures where some workers' productivity is observable. Section VII discusses the general interpretation of this model as it applies to other incentive problems. Extensions of the model are also discussed.

II. PRODUCTIVITY OBSERVABLE BY EMPLOYERS

This section derives the optimal income tax for an economy with two types of individuals: low productivity and high productivity. Employers observe productivity directly and the taxing authority observes income. Productivity is positively related to ability (which is higher for the high productivity group) and nonnegatively related to education. The inability of the taxing authority to observe ability directly and levy a lump-sum tax which depends on ability results in a second-best situation when redistributional aims are significantly large. Of course, if only efficiency is desired, it may be achieved by levying no tax at all.[2]

We now introduce some notation and assumptions that will be used throughout the paper. The low productivity group is represented by the subscript 1 and the high productivity group is represented by the subscript 2. In Section VI, when there will be two kinds of high productivity individuals, a third subscript will have to be introduced. For notational simplicity it is assumed that there are an equal number of high and low productivity workers and all variables are normalized to their per worker values. This is done without loss of generality. Individuals can produce output only when employed by firms, jobs are indivisible (i.e., there is no labor/leisure choice) Further, the productivity of an individual is independent of any action by his employer. This assumption means that returns to scale are constant in labor input and that job assignment does not affect productivity.

The following notation will be used:

y_i = education of ith type of individual

$s_i(y)$ = productivity of the ith type as a function of education

$c_i(y)$ = cost of education of the ith type as a function of education

$c_i(s)$ = cost of education required by ith type to attain a productivity of s

w_i = after-tax wage income

$N_i = w_i - c_i$ = after-tax income net of education costs

$G(N_1, N_2)$ = social welfare

All the functions given about are defined on the nonnegative reals and

all variables take on nonnegative values. The wage function (and its complement, the tax/subsidy function), however, are only defined at those levels of productivity (which is the pre-tax wage) that are observed. Alternatively, the tax/subsidy function may be thought of as being defined on all nonnegative pre-tax wages, but is nonlinear enough to induce only two (in Section III, three) pre-tax wage levels to be chosen.

Productivity functions are assumed throughout to have the following properties:

(1) $$s_2(y) > s_1(y) > 0$$

(2) and $$s_2'(y) \geq s_1'(y) \geq 0.$$

Hence, high productivity individuals have higher total and at least as high marginal productivity for a given level of education than low productivity individuals. Properties of the cost function are:

(3) $$c_1(y) \geq c_2(y) \geq 0.$$

(4) and $$c_1'(y) > c_1'(y) > 0.$$

Therefore, higher productivity workers have lower marginal and at least as low total education costs. Assumption (4) implies that (3) can hold with equalities only for $y=0$. These assumptions ensure that when productivity cannot be directly observed, more education will signal higher productivity. Productivity and cost functions are related by the second derivative condition,

(5) $$s_i''(y) - c_i''(y) < 0.$$

Further, it is assumed that at some level of education, its marginal productivity is less than its marginal cost for each type of worker. This assumption ensures that the optimal education choices are bounded.

The assumptions made about productivity and cost functions are not at all restrictive and include as a special case education having no direct contribution to productivity.

We now consider the optimal taxation problem faced by a government that wishes to maximize a concave social welfare function in an economy where firms can observe productivity and are forced by competition to pay each worker a wage (salary) equal to his productivity. To simplify the analysis it is assumed that workers can be viewed as choosing their productivity directly, which indirectly determines their education. Cost as a function of productivity is defined in terms of the education cost function by the identity

(6) $$c_i(s) \equiv c_i(y_i(s))$$

where $y_i(s) = \min y$ such that $s_i(y) \geq s$. From (1) and this definition of the cost of productivity, (3) and (4) imply, respectively,

(7a) $$c_1'(s) \geq c_2'(s) \geq 0, \quad = \text{only if } y_1(s)=0.$$

and

(7b) $$c_1(s) \geq c_2(s) \geq 0, \quad = \text{only if } y_1(s)=0.$$

The design of the optimal income tax may be thought of as the choice of the productivity for each type (s_1 and s_2) and associated after-tax wages (w_1 and w_2) that maximize social welfare. A worker with a pre-tax income of s_1 pays an income tax of s_1-w_1 (this turns out always to be negative, hence, it is a subsidy) and an worker with a pre-tax income of s_2 pays an income tax of s_2-w_2. The tax cannot be made to depend on ability; however, <u>after</u> a worker chooses his productivity level (via education) the government can infer which type of worker he is.

The choice of wages and productivities must be consistent with individual maximization of net income, which is written

$$N_i(w,s) = w - c_i(s).$$

Note that w and s are not continuously variable, they may take on only two sets of values, which are given by the tax schedule. The assumption that individuals maximize net income is equivalent to the assumption that the cost (or disutility) of education is independent of the worker's after-tax wage. This independence, though not a necessary assumption, facilitates the correction of signaling externalities by the income tax (see Section III).

The optimal taxation problem made then be written as the choice of s_1, s_2, w_1, and w_2 to

(8a) $$\text{Max} \quad G(w_1 - c_1(s_1),\ w_2 - c_2(s_2))$$

(8b) subject to $s_1 + s_2 - w_1 - w_2 \geq 0,$

(8c) $w_2 - c_2(s_2) - w_1 + c_2(s_1) \geq 0,$

(8d) and $w_1 - c_1(s_1) - w_2 + c_1(s_2) \geq 0,$

where $G_i > 0$ and G is (weakly) quasi-concave and favors equity in the following sense, for a given value of $N_1 + N_2$ welfare increases as the difference between N_1 and N_2 decreases in absolute value. (A special case of this kind of social welfare function is $G(N_1,N_2) = v(N_1)+v(N_2)$ for v concave, which is a form of the standard utilitarian social welfare function used in optimal taxation theory.) In general, this function may be viewed as representing the distributional weights of the government. Hence, the approach taken in this paper is strictly utilitarian; the relative desirability of tax and information policies depends only on the net income these policies yield individuals.

In the optimization problem above, (8b) is the balanced budget constraint (recall that there are equal numbers of each type). Constraints (8c) and (8d) are the conditions for individual optimization by 2's and 1's, respectively. This optimization is simplified considerably by the following observation:

Proposition 1: Constraint (8c) holds with equality and constraint (8d) is redundant.

Proof. Rewrite constraints (8c) and (8d) as

(9) $$\int_{s_1}^{s_2} c_1'(s)ds \geq w_2 - w_1 \geq \int_{s_1}^{s_2} c_2'(s)ds.$$

Inequalities (7a) and (9) imply that $s_2 \geq s_1$ and $w_2 \geq w_1$. Further, $N_2 \geq N_1$. This follows from rearranging (8c) and using (7b), giving

$$w_2 - c_2(s_2) \geq w_1 - c_2(s_1) \geq w_1 - c_1(s_1).$$

Suppose that s_1 and s_2 are set optimally. Then this choice of productivity levels determines the total net output, N_1+N_2. Since we know that N_2 must always be at least as great as N_1, it follows that the smaller the difference between w_2 and w_1 is made, the smaller the absolute difference between N_2 and N_1 is made. Hence, because the social welfare function favors equity, (8c) should hold with equality (making w_2-w_1 as small as possible). If (8c) holds with equality, it follows from (9) that (8d) must be satisfied.

Proposition 1 shows that the subsidy given to the low productivity workers ''tempts'' high productivity workers more than the higher productivity (and higher education costs) of high productivity coupled with the tax on their income tempts low productivity workers.

The Lagrangean for the constrained optimization (8a)-(8c) is then

(10) $$L = G(w_1 - c_1(s_1),\ w_2 - c_2(s_2)) + \lambda(s_1 + s_2 - w_1 - w_2) + \mu(w_2 - c_2(s_2) - w_1 + c_2(s_1)).$$

Noting that the first constraint must hold with equality and assuming that corner solutions are excluded,[3] the first-order conditions are

(11a) $G_1 - \lambda^* - \mu^* = 0$

(11b) $G_2 - \lambda^* + \mu^* = 0$

(11c) $-G_1 c_1'(s_1^*) + \lambda^* + \mu^* c_2'(s_1^*) = 0$

(11d) $-G_2 c_2'(s_2^*) + \lambda^* - \mu^* c_2'(s_2^*) = 0$

(11e) $s_1^* + s_2^* - w_1^* - w_2^* = 0$

(11f) $w_2^* - c_2(s_2^*) - w_1^* + c_2(s_1^*) = 0,$

where *'s denote optimal values of the variables. That the Lagrange multipliers take on positive values follows from the Kuhn-Tucker conditions for the optimization associated with the inequality constraints (8b) and (8c).

Examination of these first-order conditions gives the following result:

Proposition 2: High productivity workers choose the efficient level of education and low productivity workers choose less than the efficient level.

Proof.

(11b) and (11d) => $c_2'(s_2^*) = 1$ (efficiency).

(11a) and (11c) => $\lambda^* c_1'(s_1^*) = \lambda^* + \mu^*(c_2'(s_1^*) - c_1'(s_1^*))$.

$\lambda^* > 0$, $\mu^* > 0$, and (7a) => $c_1'(s_1^*) < 1$.

Proposition 2 illustrates the ''screening'' aspect of the optimal income tax (see Atkinson and Stiglitz, 1980). In order to identify themselves as workers who are eligible for a subsidy payment, low productivity workers must be screened out from the high productivity workers via the tax schedule. Because low productivity workers have a higher cost of acquiring productive capacity, they possess a ''comparative advantage'' in not educating themselves relative to high productivity workers. Hence, underinvestment in education serves as the screening mechanism for keeping high productivity workers from receiving the subsidy.

An inefficient choice of the education level for high productivity workers is wasteful not only in reducing the total net income of the economy, it also makes the subsidy received by low productivity workers more attractive to high productivity workers because high productivity workers will have a lower pre-tax income net of education costs than they would than when they receive the efficient level of education. That low productivity workers will prefer that high productivity workers get the efficient level of education can be easily seen algebraically. We will show that for any level of education for low productivity workers, s_1, low productivity workers will get the greatest net income, given that (8c) and (8d) must be satisified with equality, when s_2 is set at the efficient level. Solving the first constraint for w_2 and substituting into the second constraint gives

$$s_2 - c_2(s_2) = 2w_1 - s_1 - c_2(s_1).$$

So for a fixed level of s_1, w_1 is maximized when s_2 is set at the efficient level, which in turn maximizes N_1.

III. EDUCATION OBSERVABLE BY EMPLOYERS

In this section it is assumed that employers cannot observe productivity, rather, they infer it through an individual's education. As before, the taxing authority observes pre-tax wages, which in a signaling equilibrium are equal to productivity, and determines the tax/subsidy from them. This model is the two-class version of the model of optimal incentive mechanisms studied in the case of a continuum of agents by Spence and Zeckhauser (1971), Spence (1974a) and Miller (1978).

As before, employers are competitive and so profits on labor are driven to zero. Although employers do not observe productivity at the time of employment, after employees are paid firms can determine, in aggregate, the productivity of individuals with a given education. A signaling equilibrium is a wage/education schedule such that the aggregate productivity of individuals who choose a given education is equal to their (pre-tax) wage and such that no employer can offer a wage/education schedule that makes a nonzero profit given the other employers' schedules.[4] Under the assumptions made about cost and productivity functions, an equilibrium will exist and will be separating, i.e., high and low producitivity workers will choose

different education levels.[5] In such an equilibrium the pre-tax wage of each worker must be equal to his productivity; otherwise, ''cream-skimming'' would occur.

The after-tax wage of a worker of type i is, as before, w_i, and his education is y_i. Workers choose the education so as to maximize net income,

$$N_i(w,y) = w - c_i(y).$$

The only difference between the optimal income tax problem the same problem when productivity is observable by firms and when education, but not productivity, is observable by firms is that the constraints characterizing individual optimization are slightly modified. Before, a high productivity worker who considered receiving the low productivity worker's subsidy could decrease his education below that of a low productivity worker and yet have the same productivity as he and receive the same wage. When productivity is not observable, however, high productivity workers must choose the education level of a low productivity worker to receive his wage, and hence his subsidy, as wages depend only on education. As shall be demonstrated in the next section, this means that the social optimization is subject to greater constraint when productivity is observable than when it is not, so greater social welfare is possible when productivity cannot be observed.

Before making a rigorous comparison of the two cases, it is first necessary to set up the optimization problem when education is observable. In this case the optimization may be formulated as the choice of w_1, w_2, y_1, and y_2 to

(12a) $$\text{Max} \quad G(w_1 - c_1(y_1),\ w_2 - c_2(y_2))$$

(12b) $$\text{subject to} \quad s_1 + s_2 - w_1 - w_2 \geq 0,$$

(12c) $$w_2 - c_2(y_2) - w_1 + c_2(y_1) \geq 0,$$

(12d) $$\text{and} \quad w_1 - c_1(y_1) - w_2 + c_1(y_2) \geq 0.$$

The constraints have the same interpretation as in the previous model. A direct analogue of Proposition 1, i.e., constraint (12c) holds with equality and constraint (12d) is redundant, follows the proof of Proposition 1 with y_1 substituted for s_1 and y_2 substituted for s_2. In addition, it then follows that $N_2 \geq N_1$, $y_2 \geq y_1$, and $w_2 \geq w_1$.

The Lagrangean for this optimization is

(13) $$\begin{aligned} L = {} & G(w_1 - c_1(y_1),\ w_2 - c_2(y_2)) \\ & + \lambda(s_1(y_1) + s_2(y_2) - w_1 - w_2) \\ & + \mu(w_2 - c_2(y_2) - w_1 + c_2(y_1)). \end{aligned}$$

The first-order conditions are

(14a) $$G_1 - \lambda^{**} - \mu^{**} = 0$$

(14b) $$G_2 - \lambda^{**} + \mu^{**} = 0$$

(14c) $$-G_1 c_1'(y_1^{**}) + \lambda^{**} s_1'(y_1^{**}) + \mu^{**} c_2'(y_1^{**}) = 0$$

(14d) $$-G_2 c_2'(y_2^{**}) + \lambda^{**} s_2'(y_2^{**}) - \mu^{**} c_2'(y_2^{**}) = 0$$

(14e) $$s_1(y_1^{**}) + s_2(y_2^{**}) - w_1^{**} - w_2^{**} = 0$$

(14f) $$w_2^{**} - c_2(y_2^{**}) - w_1^{**} + c_2(y_1^{**}) = 0.$$

**'s indicate the optimal choices.

Just as Proposition 1 held for this case, so will Proposition 2. This can be proven from the first-order conditions (14a)-(14d) as follows:

(14b) and (14d) => $s_2'(y_2^{**}) - c_2'(y_2^{**}) = 0$ (efficiency for 2's).

(14a) and (14c) => $$s_1'(y_1^{**}) - c_1'(y_1^{**}) = \frac{\mu^{**}(c_1'(y_1^{**}) - c_2'(y_1^{**}))}{\lambda^{**}} > 0.$$

Hence, y_1^{**} is, by the second derivative condition, (5), less than the efficient level.

In the presence of optimal income taxation, the informational externalities that result from employers being unable to observe productivity are, in terms of efficiency, fully corrected. The incentive effects of the optimal tax become qualitatively the same as those when productivity is observable: only underinvestment in education is possible. This result corresponds to the result in Miller (1978) that in economies with a continuum of agents the disincentive effect of the income tax will always dominate the signaling effect when the tax is optimally set and the social welfare functional is concave.

IV. A COMPARISON OF THE TWO INFORMATION STRUCTURES

Before examining information structures that are intermediate between the ones discussed in the previous two sections, we will show that if these structures are the only possible, then having productivity observed directly will never result in greater welfare than having productivity signaled. The intuitive reason for this result is that the signaling information structure has both efficiency and distributional advantages. The efficiency advantage results from the fact that the signaling effect offsets, but does not dominate, the disincentive effect of the tax. The distributional advantage is rooted in the fact that denying high productivity workers the full exploitation of their competitive advantage by forcing them to signal that advantage serves, in essence, to redistribute income from them to the low productivity workers. Another way of considering the redistributional aspect of signaling is that the condition on individual income maximization in the signaling case (12c) is less constraining (more income may be redistributed) than the corresponding constraint when productivity is observed (8c). The following theorem is proved by utilizing that property of the contraints.

<u>Theorem</u> <u>1</u>:

$$G(w_1^{**}-c_1(y_1^{**}),w_2^{**}-c_2(y_2^{**}))\geq G(w_1^{*}-c_1(s_1^{*}),w_2^{*}-c_2(s_2^{*})).$$

Proof. Let $y_i^* \equiv y_i(s_i^*)$. We show that w_1^*, w_2^*, y_2^*, and y_2^* satisfy (12b) and (12c). (12b) holds trivially. By (6) and (8c),

$$w_2^* - c_2(y_2^*) - w_1^* + c_2(y_2(s_1^*)) \geq 0. \tag{15}$$

From (1), $y_2(s_1^*) \leq y_1(s_1^*) = y_1^*$ and by (4),

$$c_2(y_2(s_1^*)) \leq c_2(y_1^*). \tag{16}$$

Combining (15) and (16) give

$$w_2^* - c_2(y_2^*) - w_1^* + c_2(y_1^*) \geq 0. \tag{17}$$

Hence, (12c) holds for w_1^*, w_2^*, y_1^*, and y_2^* and the theorem follows immediately.

Note that the solution to (8a)-(8c) may not satisfy the redundant constraint, (12d). This constraint is redundant only at an optimum; a suboptimal allocation which satisfies (12c) need not satisfy (12d). This is because in the signaling case it is easier for low productivity workers to receive the high productivity wage; they need only signal that they are high productivity workers rather than achieve the productivity of high productivity workers. Because the optimal allocation when productivity is directly observable may not be feasible under signaling (but some feasible allocation will yield more welfare), the signaling economy should not be expected to yield a Pareto-superior allocation.

V. A NUMERICAL EXAMPLE

This section gives a numerical example of the optimizations described in the previous sections. The cost and production functions are taken to be

$$c_1(y) = 2y^2, \quad c_2(y) = y^2, \quad s_1(y) = 8y, \text{ and } \quad s_2(y) = 16y.$$

$$\Rightarrow c_1(s) = \frac{s^2}{32} \text{ and } c_2(s) = \frac{s^2}{256}.$$

The social welfare is taken to be multiplicative (Cobb-Douglas),

$$G(N_1, N_2) = N_1 N_2.$$

The efficient levels of productivity and education are:

$$s_1^{eff} = 16 \qquad s_2^{eff} = 128$$

$$y_1^{eff} = 2 \qquad y_2^{eff} = 8.$$

We now set up the Lagrangean and first-order conditions when productivity is observable and find the optimal values of all relevant variables.

$$L = (w_1 - \frac{s_1^2}{32})(w_2 - \frac{s_2^2}{256}) + \lambda(s_1+s_2-w_1-w_2) + \mu \; (w_1 - \frac{s_2^2}{256} - w_2 + \frac{s_1^2}{256}) = 0$$

(18a) $$\frac{\partial L}{\partial w_1} = w_2^* - \frac{(s_2^*)}{256} - \lambda^* - \mu^* = 0$$

(18b) $$\frac{\partial L}{\partial w_2} = w_1^* - \frac{(s_1^*)^2}{32} - \lambda^* + \mu^* = 0$$

(18c) $$\frac{\partial L}{\partial s_1} = \frac{-s_1^*}{16} \; (w_2^* - \frac{(s_2^*)^2}{256}) + \lambda^* + \mu^* \frac{s_1^*}{128} = 0$$

(18d) $$\frac{\partial L}{\partial s_2} = \frac{-s_2^*}{128} \; (w_1^* - \frac{(s_1^*)^2}{32}) + \lambda^* - \mu \frac{s_1^*}{128} = 0$$

(18e) $$\frac{\partial L}{\partial \lambda} = s_1^* + s_2^* - w_1^* - w_2^* = 0$$

(18f) $$\frac{\partial L}{\partial \mu} = w_2^* - \frac{(s_2^*)}{256} - w_1^* + \frac{(s_1^*)^2}{256} = 0$$

(18b) and (18d) $\Rightarrow s_2^* = 128 \Rightarrow y_2^* = 8 = y_2^{eff}$

Add (18a) to (18b) to get:

(19) $$\lambda^* = \frac{w_1^* + w_2^*}{2} - \frac{(s_1^*)^2}{64} - 32.$$

Subtract (18a) from (18b) to get:

(20) $$\mu^* = \frac{w_2^* - w_1^*}{2} + \frac{(s_1^*)^2}{64} - 32.$$

Add (18c) to (18f) to get:

(21) $$w_2^* = \frac{s_1^*}{2} - \frac{(s_1^*)^2}{512} + 96.$$

Substituting (18f) into (19) and (18c) into (20),

(22) $$\lambda^* = \frac{s_1^*}{2} - \frac{(s_1^*)^2}{64} + 32$$

$$\mu^* = \frac{7(s_1^*)^2}{512}. \qquad (23)$$

Then substitute (21), (22), and (23) into (18c) and simplify

$$\frac{15(s_1^*)^3}{65536} - \frac{3(s_1^*)^2}{64} - \frac{3s_1^*}{2} + 32 = 0. \qquad (24)$$

The relevant root is

$$s_1^* \;\; 14.9001 \Rightarrow y_1^* \;\; 1.8625 < y_1^{eff}.$$

The solution is then:

$$w_1^* \approx 39.8837 \qquad w_2^* \approx 103.0164$$

$$N_1^* \approx 32.9457 \qquad N_2^* \approx 39.0164$$

$$G(N_1^*, N_2^*) \approx 1285.4261$$

$$N_1^* + N_2^* \approx 71.9622$$

The net output of the economy, $N_1^* + N_2^*$, is very close to the maximum possible net output of 72. The value for G is also not far from its first-best value, 1296. Of course, with a social welfare function that placed more weight on distributional considerations than does the multiplicative form, the deviation from the first-best in both efficiency and welfare terms would be greater. When productivity is signaled the economy does even better; greater welfare is achieved with less loss in efficiency.

When productivity is signaled the Lagrangean and first-order conditions are:

$$L = (w_1 - 2y_1^2)(w_2 - y_2^2) + \lambda(8y_1 + 16y_2 - w_1 - w_2) + \mu(w_2 - w_1 + y_1^2 - y_2^2)$$

$$\frac{\partial L}{\partial w_1} = w_2^{**} - (y_2^{**})^2 - \lambda^{**} - \mu^{**} = 0 \qquad (25a)$$

$$\frac{\partial L}{\partial w_2} = w_1^{**} - 2(y_1^{**})^2 - \lambda^{**} + \mu^{**} = 0 \qquad (25b)$$

$$\frac{\partial L}{\partial y_1} = -4y_1^{**}(w_2^{**} - (y_2^{**})^2) + 8\lambda^{**} + 2\mu^{**}y_1^{**} = 0 \qquad (25c)$$

$$\frac{\partial L}{\partial y_2} = -2y_2^{**}(w_1^{**} - 2(y_1^{**})^2) + 16\lambda^{**} - 2\mu y_2^{**} = 0 \qquad (25d)$$

$$\frac{\partial L}{\partial \lambda} = 8y_1^{**} + 16y_2^{**} - w_1^{**} - w_2^{**} = 0 \qquad (25e)$$

$$\frac{\partial L}{\partial \mu} = w_2^{**} - w_1^{**} + (y_1^{**})^2 - (y_2^{**})^2 = 0 \qquad (25f)$$

The first-order conditions are solved the same way as before, yielding

$$y_1^{**} \approx 1.9486 < y_1^{eff} \qquad y_2^{**} = 8. = y_2^{eff}$$

$$w_1^{**} \approx 41.6930 \qquad w_2^{**} \approx 101.8959$$

$$N_1^{**} \approx 34.0988 \qquad N_2^{**} \approx 37.8959$$

Comparing the two models gives

$$G(N_1^{**}, N_2^{**}) \approx 1292.2054 > G(N_1^*, N_2^*)$$

and

$$N_1^{**} + N_2^{**} \approx 71.9947 > N_1^* + N_2^*$$

It is seen then that with signaling the efficiency loss is very slight and that it is possible to redistribute somewhat more than one additional unit of income from the high productivity workers to the low productivity workers. This example is, of course, just an illustration of the preceding theorem and the numbers chosen have no significance other than that they make the numerical derivation of the optimum tractable. Although the improvement in welfare in this example was miniscule, in general, the size of the improvement will depend on the social welfare function and the differences in the cost and productivity functions of the two groups. The larger the differences in the cost and productivity functions between the two groups, the easier it will be to satisfy the individual optimization constraint. Also, the relative sizes of the two groups (for the example they were taken to be equal) will also affect the welfare gain from signaling.

VI. A THEOREM ON MORE GENERAL INFORMATION STRUCTURES

This section considers information structures that are intermediate in information content between productivity being observed for all workers and being signaling by all workers. Such an information structure will be characterized by the parameter p, where $p \varepsilon [0,1]$ is the proportion of (high productivity) workers whose

productivity is <u>not</u> observable. Since low productivity workers essentially receive their jobs by default, it does not matter whether or not their productivity is observable. There are at least two ways in which the information structure may be interpreted. One is that it may be considered as rules that restrict the transmission of information that accurately conveys worker productivity. Alternatively, these rules may require that a certain proportion of hiring be based on an ''objective'' signaling criterion. Affirmative action programs, although intended to remedy discrimination, would appear to be a government action that works in this way. The strictness of the such hiring rules and the level of their enforcement would determine the value of p, either in a deterministic or probabilistic way. Note that in the context of this model, aside from the impact they have on discrimination, affirmative action rules are redistributive in nature.

It is proved that even when these mixed information structures are considered, there is still no information structure that yields more welfare than the one where no worker's producitivity is observable, (p=1). Intuitively, this means that as far as the structure of incentives is concerned, there is no favorable interaction between the two kinds of high productivity workers. All that occurs is that as p increases, the constraints on the economy are diminished. This point is illustrated by the construction of an auxiliary optimization problem which embodies constraints (8c) and (12c). This optimization problem may be written as:

$$\text{(26a)} \quad \text{Max} \quad G(w_1 - c_1(s_1),\ pf(w_2 - c_2(s_2)) + (1 - p)f(w_3 - c_3(s_3)))$$

$$\text{(26b)} \quad \text{subject to} \quad s_1 + ps_2 + (1 - p)s_3 - w_1 - pw_2 - (1 - p)w_3 \geq 0$$

$$\text{(26c)} \quad w_3 - c_3(s_3) - w_1 + c_3(s_1) \geq 0$$

$$\text{(26d)} \quad \text{and} \quad w_2 - c_2(s_2) - w_1 + c_2(y_1(s_1)) \geq 0 \ .$$

The high productivity workers whose productivity is not observable are denoted by 2's and those whose productivity is observable are denoted by 3's. The maximand, (26a) has been altered to reflect the fact that high productivity workers are now of two types; the function f, which is taken to be (weakly) concave, gives the distributional weights attached to the two kinds of high productivity workers. Inequality (26b) is the budget constraint; (26c) and (26d) are the individuals optimization constraints. As before, the redundant constraints, which are still redundant, may be omitted. For $p=0$, this optimization is equivalent to (8a)-(8c) and for $p=1$, it is equivalent to (12a)-(12c). Note that it is less constrained that the true optimization problem for the mixed case because it does not include the restriction that the tax system levies the same tax on individuals with the same income ($s_2=s_3 \Rightarrow w_2=w_3$). Also, individual optimization constraints between 2's and 3's are omitted. These omissions are inconsequential if the economy is centralized, the employer and the taxing authority are the same entity. In this case this optimization problem is the proper one. The Lagrangean for this problem is

$$\text{(27)} \quad L = G(w_1 - c_1(s_1),\ pf(w_2 - c_2(s_2)) + (1 - p)f(w_3 - c_3(s_3)))$$

$$+ \lambda(s_1 + ps_2 + (1 - p)s_3 - w_1 - pw_2 - (1 - p)w_3)$$

$$+ \mu(w_3 - c_3(s_3) - w_1 + c_3(s_1))$$

$$+ \rho(w_2 - c_2(s_2) - w_1 + c_2(y_1(s_1)))$$

The first-order conditions are then

$$\text{(28a)} \quad G_1 - \lambda - \mu - \rho = 0.$$

$$\text{(28b)} \quad G_2 pf'(N_2) - \lambda p + \rho = 0$$

$$\text{(28c)} \quad G_2(1 - p)f'(N_3) - \lambda(1-p) + \mu = 0$$

$$\text{(28d)} \quad -G_1 c_1'(s_1) + \lambda + \mu c_3'(s_1) + \rho c_2'(y_1)y_1'(s_1) = 0$$

$$\text{(28e)} \quad -pG_2 f'(N_2)c_2'(s_2) + \lambda p - \rho c_2'(s_2) = 0$$

$$\text{(28f)} \quad -(1 - p)G_2 f'(N_3)c_3'(s_3) + \lambda(1 - p) - \mu c_3'(s_3) = 0.$$

In addition, there are the usual first-order conditions associated with the constraints. It is understood that the variables in (28a)-(28f) represent their values at an optimum.

To show that $p=1$ gives the optimal information structure, we show that for the auxiliary optimization welfare never decreases as p increases.

<u>Lemma</u>.

$$\frac{dG}{dp} \geq 0.$$

Proof.

By the envelope theorem,

(29) $$\frac{dG}{dp} = G_2(f(N_2) - f(N_3)) + \lambda(s_2 - s_3 - w_2 + w_3) .$$

But,

$$(28b) \text{ and } (28e) \Rightarrow c_2'(s_2) = 1$$

$$(28c) \text{ and } (28f) \Rightarrow c_3'(s_3) = 1.$$

But 2's and 3's have the same cost function, so $s_2 = s_3$. Then (29) reduces to

(30) $$\frac{dG}{dp} = G_2(f(N_2) - f(N_3)) + \lambda(w_3 - w_2)$$

when $w_2 = w_3$, the lemma follows immediately. If $w_2 \neq w_3$, (26c) and (26d) imply that $w_2 < w_3$. The concavity of f gives

(31) $$f(N_2) - f(N_3) \geq -f'(N_2)(w_3 - w_2)$$

Then (30) and (31) $\Rightarrow \frac{dG}{dp} \geq (-G_2 f'(N_2) + \lambda)(w_3 - w_2)$.

From (28b),

$$-G_2 f'(N_2) + \lambda = \frac{\rho}{p} \geq 0,$$

proving the lemma.

The feasible set described by (28b)-(28d) is identical with

that of the true optimization problem for p=0 and p=1 and is at least as large for intermediate values of p. Hence, we have

Theorem 2: The optimal choice of p is p = 1.

Proof. Immediate from the Lemma.

VII. CONCLUDING REMARKS

Although this model was developed in terms of the education signaling model, it is applicable in other situations where effort affects productivity. The rat-race model of assembly-line speed and the labor/leisure model of worker behavior under income taxation fall into this class. As long as the workers' utility functions are separable between after-tax wage income and effort, the results are unchanged. In a centralized economy the results may be interpreted as saying that wages should not depend solely on output, that considering just individual effort, if observable, leads to higher welfare.

Various extensions of the model may prove fruitful. There is nothing special, it would appear, about the assumption of two kinds of workers; the effects that lead to higher welfare under signaling are still present with more types of workers, including a continuum. Relaxing the restrictions on the functional form of worker's utility is also of interest, because in that case the income tax is not equivalent to an education tax, so more complex incentive structures are possible. In this setting it is also possible that the results of this paper no longer hold as strongly.

Further generalization of the model are possible by introduced noise into the system, so that the choice of information structure includes the choice of a signaling mechanism from a number of noisy signals. Work by Spence (1976b) indicates that in the absence of corrective taxation, such noise is desirable. Noise may also be present in the economy in that productivity may only be imperfectly observable, which will make the individual optimization constraints less binding at some cost in efficiency.

FOOTNOTES

I would like to thank Mary O'Keeffe, Louis Wilde, Edward Green and members of the Caltech Theory Workshop for their useful comments.

1. See Spence (1974b) for a general discussion of signaling behavior.

2. Of course, if the government wishes to raise revenues, a lump-sum tax independent of abilities is efficient.

3. Since both groups are given wages that are greater than s_1, which is positive, they can never be zero. If the marginal productivity of education to low productivity workers is low enough (e.g., always zero), the productivity corresponding to zero education will be chosen and the corresponding first-order condition will not hold with equality. This will not affect any of the analysis except for the fact that if education adds no productivity, it cannot be underinvested in. If the signal is also unproductive for high productivity workers, they still end up at efficient investment (zero).

4. See Spence (1976a) for a discussion of this signaling equilibrium.

5. Unless, of course, education is productive for neither group. The resulting allocation is a constrained Pareto optimum; hence, by the results of Wilson (1977) there is no problem with the existence of the equilibrium. In essence, the tax system provides the cross-subsidies necessary to maintain an equilibrium in the face of competition.

REFERENCES

Arrow, K. J. ''Higher Education as a Filter.'' Journal of Public Economics 2 (1973):193-216.

Atkinson, A. B. and J. E. Stiglitz. Lectures on Public Economics. New York: McGraw-Hill Book Company (1980).

Green, J. ''Statistical Decision Theory Requiring Incentives for Information Transfer.'' Harvard Institute of Economic Research Discussion Paper #708 (1979).

Green J. and N. Stokey. ''The Value of Information in the Delegation Problem.'' Harvard Institute of Economic Research Discussion Paper #776 (1980)

Hirshleifer, J. ''The Private and Social Value of Information and the Reward to Incentive Activity.'' American Economic Review 61 (1971):561-574.

Holmstrom, B. ''Moral Hazard and Observability.'' Bell Journal of Economics 10 (1979):74-91.

Hylland, A. and R. M. Zeckhauser. ''Distributional Objectives Should Affect Taxes but not Program Choice or Design.'' Scandanavian Journal of Economics 81 (1979):264-84.

Miller R. ''Signaling and Optimal Income Taxation.'' mimeo., Harvard University, (1978).

Mirrlees, J. ''An Exploration in the Theory of Optimum Taxation.'' Review of Economic Studies 38 (1971):175-208.

Phelps, E. ''Taxation of Wage Income for Economic Justice.'' Quarterly Journal of Economics 87 (1973):331-354.

Shavell, S. ''Risk Sharing and Incentives in the Principal and Agent Relationship.'' Bell Journal of Economics 10 (1979):53-73.

Shavell, S. ''A Note on Efficiency Vs. Distributional Equity: Should Distributional Equity Matter Given Optimal Income Taxation.'' American Economic Association Papers and Proceedings 71 (1981):414-418.

Spence, M. ''Competitive and Optimal Responses to Signals: An Analysis of Efficiency and Distribution.'' Journal of Economic Theory 7 (1974a):296-332.

Spence, M. Market Signalling. Cambridge: Harvard University Press, 1974b.

Spence, M. ''Competition in Salaries, Credentials, and Signalling Prerequisites for Jobs.'' Quarterly Journal of Economics 90 (1976):51-75.

Spence, M. ''Signalling and Screening.'' Harvard Institute of Economic Research Discussion Paper #467, (1976b).

Wilson, C. ''A Model of Insurance Markets with Incomplete Information.'' Journal of Economic Theory 16 (1977): 167-207.