Incentives, risk, and information: notes towards a theory of hierarchy

Joseph E. Stiglitz

Professor of Economics
Stanford University

This paper analyzes the role of incentives, risk, and information in determining the structure of employment contracts. In particular, we focus on the functions performed by piece rate versus time rate payment systems and by supervisors. The relative reliance on piece rates versus time rates is related to risk sharing, to the use of the payment system as a method of screening employees, and to differential information concerning the difficulties of the tasks being performed. The choice of payment system thus depends on the attitudes toward risk of workers and employers, effort supply elasticities, the sources and magnitude of the uncertainties, and the nature of the supervision used in the employment relation. The supervisor is viewed as monitoring inputs (enforcing contracts), screening individuals, obtaining information about the state of the world, etc. These roles are related to the nonconvexities associated with information.

1. Introduction

■ There is an enormous variety of contractual arrangements under which workers sell their services to firms. Economic theory has had little to offer, however, by way of explaining why particular firms choose particular contractual arrangements. The “details” are swept aside with the assertion that, in competitive markets, workers will get paid their marginal product. If there is some uncertainty, workers sell their labor, contingent upon the state of nature.

The object of this paper is to analyze some of the determinants of the particular structure of contracts chosen by particular firms. Since contracts differ with respect to a large number of provisions—the amount of guaranteed pay, the incentive pay,

Joseph E. Stiglitz received the B.A. from Amherst College in 1964 and the Ph.D. from the Massachusetts Institute of Technology in 1966. His research centers on the economics of information and uncertainty.

This is a much revised version of a paper originally presented at the Berlin Symposium on Planning, 1973. A revised version was presented at the Symposium on the Economics of Internal Organization which was held at the University of Pennsylvania in September 1974. The Symposium was supported by grants from the National Science Foundation (NSF-GS-35889X) and the General Electric Foundation. The author is indebted to Lars Osberg, A. B. Atkinson, A. M. Spence, and the participants in the symposium for their helpful comments. He is particularly indebted to John Pencavel for his insightful comments. The results reported in “An Essay on the Economics of Work Effort and Wage Payments Systems” (Stanford, mimeo, 1974) closely parallel those discussed here. This work was supported by National Science Foundation Grant SOC 74-22182 at the Institute for Mathematical Studies in the Social Sciences, Stanford University.

the rules by which the incentive is adjusted, the degree of supervision, promotion policy, retirement policy, etc.—it is not surprising that a larger number of detailed characteristics of individuals and production processes are relevant in determining the exact structure of contracts chosen. Theoretical work in this area is almost of necessity somewhat messy. We shall accordingly focus our analysis on only two aspects of the job contract: the choice of payment by time versus piece rates, and the function of supervision (monitoring). These two aspects are in fact closely related: the former can be viewed as an analysis of equilibrium wage contracts when certain kinds of monitoring are prohibitively expensive. Even in our analysis of these two aspects of the job contract, we shall have to narrow our discussion to a few of the primary determinants of the equilibrium wage contract within a competitive labor market.

As the title of the paper suggests, the problems we focus on are those related to information, risk, and incentives. In fact, the problems of risk and incentives can be viewed as aspects of the economics of imperfect information: risk reflects imperfect information about states of nature (rainfall), individuals' abilities, the tasks to which an individual will be assigned, etc. Conventional risk sharing arrangements face difficulties because, to the extent that there is risk sharing, there is likely to be a reduction in incentives (known as moral hazard in the insurance literature); but incentives are a problem mainly because of difficulties in monitoring individuals' inputs and outputs.

Almost all labor contracts allow both the employer and the employee a certain amount of discretion. They do not specify precisely what each party will do in each state of nature. The employer is allowed to assign the workers different tasks;¹ the worker is allowed some latitude in how he performs the tasks. This discretionary action is itself a consequence of imperfect information (it is impossible for the firm to predict the profit maximizing set of tasks to which workers should be assigned). Even when the contract specifies, for example, the task that is to be assigned in a particular state of nature, the employee may find it difficult to ascertain whether that state of nature has occurred, i.e., whether the firm is complying with the terms of the contract; similarly, the worker may agree to supply a given level of effort, but the firm may find it difficult to know whether an observed low output is a result of low effort or of some exogenous event.

This paper is concerned with the lack of information of the employer and the discretionary actions of employees. The symmetric problem involving the discretionary action of the employee is in one sense easier, in another more difficult. Firms as permanent institutions acquire a reputation; thus what individuals are concerned with is distribution of tasks which are actually assigned. We assume a competitive labor market, and the terms at which a firm can attract laborers depends on this reputation which it has established. This is a long-run considera-

¹ This is essentially the central point of Simon's classic paper on the employment relationship.

tion; in the short run compliance with the implicit terms of the contract may be obtained only through the threat of the individual's quitting (and legal and union action). The analysis of this side of the market would, however, take us beyond the scope of this paper.

□ **A general model.** Because the problems of determining the equilibrium wage contract are extremely complicated, we consider a number of related, simplified models. But before beginning the analysis, we shall present a general (but still not the most general) model; the subsequent analysis can be viewed as special cases of this model.

Firms attempt to find that contract which will maximize profits, subject to the constraint that some workers accept the contract. The various provisions of the contract determine not only who applies and accepts employment at a given firm, but also performance on the job. Thus, we replace the competitive wage-taking hypothesis with a "utility-taking" hypothesis.² Firms are sufficiently small so that they believe they face horizontal supply schedules of laborers of different character, i.e., for any contract they offer, they face essentially either a zero supply or an indeterminantly large supply. The difficult question is, which of the infinite number of possible contracts maximizes profits.

More formally, we assume there are a large number of tasks within the firm. We can describe the output of the j th individual assigned to the i th task by the k th firm by

$${}_kQ_i^j = {}_kF_i^j(e_i^j, \theta^j, \theta_i, \theta_i^j). \quad (1)$$

The output (per unit of time) is a function of the effort, e_i^j , supplied by the worker (per unit of time) and a vector of other variables. These may relate to (a) some general characteristics of the individual (his IQ), θ^j ; (b) some parameter of the technology, θ_i (e.g., the "difficulty" of the task or the amount of rainfall); and (c) some parameters describing characteristics of the individual which relate specifically to his performance of this task, θ_i^j .

The effort of the individual is, in turn, a function of the contract between the firm and the given individual which we denote for the moment by ω^j , his information structure (e.g., his knowledge concerning his own abilities and the technological parameters of the task to which he has been assigned), denoted by Ω^j , and his utility function.

² The reason that we do not replace the wage-taking hypothesis with a more general contract-taking hypothesis is that there are clearly an infinite number of possible contracts, only a few of these will be available on the market at any time, and there will in general not exist implicit prices by which various contemplated provisions may be evaluated. Thus, the firm must consider not only what is the best contract among those which are presently offered, but the best among all those which could be offered.

This kind of framework for analyzing equilibrium contracts was originally explored in the context of an agricultural share-cropping economy, in Stiglitz [32]. There, a contract specified more than just a wage; it entailed a fixed payment, a share, the amount of land the individual was to work on, and if it is observable, the amount of labor input.

For simplicity, we shall assume that the individual's information structure is not affected by his actions (other than acceptance of a job) but may be a function of time. Given the firm's wage contract, and the set of other available wage contracts, the individual is assumed to solve the dynamic programming problem of maximizing his lifetime expected utility. This will determine his level of effort at each date that he is on the job:

$$e_i^j(t) = e_i^j(\omega^j, \Omega^j(t)). \quad (2)$$

The firm is concerned with maximizing its expected profits.³ Expected returns R are a function of the outputs of the various tasks,

$$R_k = R_k(\mathbf{Q}_i), \quad (3)$$

where⁴

$${}_k\mathbf{Q}_i = \sum_k \mathbf{Q}_i^j, \quad (4)$$

where the summation is taken over all the individuals assigned to the task.

Labor costs are a function of the various contracts offered and actually accepted, of the quality of people accepting employment, and of the effort which they provide. The firm may offer the individual a choice among several contracts. The set of wage contracts offered to an individual will be a function of the characteristics of the individual which are observable at that time, i.e., if ω_k^{lj} is the l th contract offered the j th individual by the k th firm,

$$\omega_k^{lj} \in W_k(\hat{\theta}^j), \quad (5)$$

where $\hat{\theta}^j$ are characteristics which are observable. An individual with information structure Ω^j would accept a job at the given firm if and only if his lifetime expected utility on that job (given his information) exceeded that of any other firm. Let $W^j(\omega_k^{lj})$ be the expected utility of the individual with contract l at firm k , then

$$W^j(\omega_k^{lj}) \geq W^j(\omega_n^{mj}) \quad (6)$$

for all $\omega_n^{mj} \in \omega_n(\hat{\theta}^j)$, all m, n . This determines those who apply to the firm and the contracts (jobs) for which they apply. Individuals who have the same observable characteristics may have other characteristics which are unobservable but important for production; some of these characteristics can be inferred from the jobs (contracts) for which the individual applies. We assume that if the firm hires workers under a given contract, it obtains a random sample of those who have applied for the job. Thus, if the average quality of those who apply for one contract is higher than that for another, we assume, on average, the quality of those who are hired under one contract is higher than that under another.

³ The modifications required for risk averse firms are obvious.

⁴ The restriction embedded in (4) is unnecessary (a general functional relationship would do as well).

Thus, we can write the expected profits of the firm simply as a function of the contracts it offers:⁵

$$\pi = \pi(W_k(\hat{\theta})). \quad (7)$$

If $W_k^*(\hat{\theta})$ denotes the set of contracts for which some individuals will accept employment at the given firm, then, simply put, the problem of the firm is to

$$\begin{aligned} &\max \pi(W_k(\hat{\theta})) \\ &\text{subject to } W_k \in W_k^*(\theta) \end{aligned} \quad (8)$$

and the problem of this paper is to characterize the solutions to (8).

In much of our analysis, we further assume free entry, so that in market equilibrium, the maximized value of profits are zero.

□ **Some general implications of the general model: on the relationship between internal and external labor markets.** It is the underlying hypothesis of this paper that insight into the determinants of the kinds of contracts in use can be derived by analyzing the solutions to the competitive market equilibrium problem posed here. Obviously, the full solution to the maximization problems involved in the analysis requires information and computations beyond the capabilities of many participants in the market; still, if there are large and significant advantages of one contractual arrangement over another, firms that “discover” the preferred contractual arrangement will find they can increase profits and the particular contractual arrangement will be imitated. Thus, it might be argued that there is an evolutionary tendency of the economy to gravitate to the contractual arrangements analyzed here. At the same time, considerations of “bounded rationality” do suggest that we focus our attention on firm choices over relatively simple contractual arrangements.⁶

This general framework makes clear that the dichotomy that has sometimes been suggested between internal and external labor markets is not a completely accurate one. The characteristics of the internal labor market are determined largely by the external labor market, in two senses: the terms of the structure of the labor market affect the wages at which it can attract labor and the kinds of laborers it can attract; and secondly, the kinds of contracts that are feasible depend on subsequent employment opportunities elsewhere, e.g., if information about individual qualities can be ascertained from the jobs to which they are assigned, and long-term contracts are not enforceable, in later

⁵ Obviously, there are a number of other important variables, such as the stock of capital, which affect the profitability of the firm.

⁶ It should, however, be noted that the kind of imperfect information involved in the argument for bounded rationality is of a different sort from those on which this analysis is focused. The latter is concerned with the unobservability of particular characteristics of particular individuals; the former with lack of knowledge concerning the distribution of characteristics (tastes, abilities, opportunities) in the population as a whole and with the limitations on the ability to solve precisely complicated maximization problems.

stages of individuals' lives, they will receive a pay commensurate with their abilities.

In the subsequent sections, we shall specialize this model in various ways. In Section 2 we consider only labor contracts which depend on the output of the individual and his time (assumed to be the only observables). We assume that all individuals are identical, and the only random parameter is θ_i ,⁷ a parameter of the technology. The individual will be assigned only one task. We next consider the case where, in addition, individuals have different abilities (θ^i) and those more able to know that they are more able; then the set of people who accept employment at any wage offer becomes important (i.e., relations (6) now play a crucial role). The concluding parts of Section 2 consider the case where there is a single ability, but a distribution of tasks to be completed.

Section 3 shows how the introduction of supervisors can expand the set of contracts which can be offered, e.g., payments may be made functions of variables other than time or output; and rather than the same contract's being offered to everyone, a different contract is offered depending on the ability of the individual.

■ It is a characteristic of many, if not most, job contracts that payment is related both to the time spent on the job and to performance on the job. The former we refer to as a time rate. Reward for performance may take either the form of higher pay in the given period (which we refer to as a piece rate) or a higher probability of promotion (a higher probability of a higher wage at some date in the future). Our objective here is the determination of the relative importance of time and piece rates. Moreover, we focus here on the case where there is no supervision, so the level of effort of the individual cannot be ascertained and neither his ability nor any of the technological parameters which are assumed initially to be unknown can be directly observed. Since we are considering a choice between piece rates and time rates, we obviously assume that both the labor time provided by the individual and the *physical output* of the individual can be identified.

Both the compensation for time and performance are generally nonlinear functions; e.g., individuals usually receive higher time wages for working more than 40 hours a week, and piece rates systems often have a minimum quota, so that compensation as a function of performance appears as in Figure 1b. Indeed, the pure quota system, in which individuals are paid a fixed amount if they meet their quota, and nothing more if they exceed it, can be viewed as a limiting form of a piece rate system (see Figure 1c). Throughout, we shall only consider linear time rate systems⁸ and we shall pay particular attention to

2. Piece rate versus time rate

⁷ There is an alternative interpretation which in individuals may differ, but they are uninformed about their abilities.

⁸ Since the major reasons generating nonlinear time rates are quite different from the kind of considerations we focus on in this paper.

FIGURE 1a

LINEAR PIECE RATE SYSTEM

FIGURE 1b

BASE PAY WITH QUOTA

FIGURE 1c

PURE QUOTA SYSTEM

linear piece rate systems, both because they are analytically simple and because they appear to be important in practice.

We can identify several factors affecting the relative reliance which will be placed on piece rates versus time rates:

(1) Because the variability of the workers' income will increase with a greater reliance on piece rates, the less the risk and the smaller the degree of risk aversion, the greater the reliance on piece rates.⁹

(2) Because piece rates reward speed, they tend not to provide the correct incentives for quality or proper care of equipment;¹⁰ thus, the more important are the variables other than speed of performance which are under the control of workers and the more difficult it is to observe these, the less the reliance on piece rates.

(3) Because one of the difficulties in operating a piece rate system is the determination of the appropriate piece rate, the greater the frequency of a change in technology, and the greater the difficulty in ascertaining the difficulty associated with any task (or in determining the value added involved in any single operation), the less the reliance on piece rates.

(4) The greater the responsiveness of individuals to monetary incentives, the greater the reliance on piece rates.

(5) Because workers with greater ability receive a higher pay under piece rate systems, the greater the heterogeneity of the labor force with respect to its ability to perform the required tasks and the greater the difficulties associated with ascertaining these abilities directly, the greater the reliance on piece rates.

In the subsequent subsections, we attempt to derive, for some special cases, the precise determinants of the equilibrium contract.

□ **Linear contracts with risk averse workers.** Let θ be the risk variable representing any factor affecting the relationship between effort and output. We let e be the work effort of the individual. The value of output per unit of time is assumed to be proportional to the product of θ and e , and we choose our units so that the proportionality constant is unity. (For simplicity, no other factors are employed so that output and marginal productivity are identical.) A linear incentive schedule thus pays an individual a time rate I , plus a piece rate r . The individual's income, C (per unit of time), is then

⁹ Obviously, reducing the risk faced by the worker increases the risk faced by the firm. What are crucial are differences in risk aversion. We assume, however, that the firm is risk neutral, so the contract depends on the risk aversion of workers. For a more general analysis, in a somewhat different context, see Stiglitz [32].

¹⁰ If the employee is required to pay for his own machines, then this is no longer a problem. Thus, the magnitude of this problem depends on how we organize the production process. A similar observation applies to agriculture: if we hired workers to pick weeds, and paid them by the weight of weeds picked, they would pick the largest weeds, not necessarily the ones which would most likely interfere with the growth of the vegetables; if we hire workers to grow vegetables, and pay them in proportion to their output of vegetables, they will pick the correct weeds. See Stiglitz [32].

$$C = I + r\theta e. \quad (9)$$

But in competitive equilibrium (letting superbars denote means), with free entry,

$$\overline{C} = I + r\overline{\theta e} = \overline{\theta e}. \quad (10)$$

Thus

$$C = r\theta e + (1 - r)\overline{\theta e} = r(\theta e - \overline{\theta e}) + \overline{\theta e}. \quad (11)$$

The individual's income is a weighted average between the mean marginal product of those working under the given contract and the individual's own marginal product. Alternatively the individual's income is equal to the mean marginal product plus an incentive pay equal to the difference between his labor services and that of the mean.

If all individuals were identical, the contract chosen would be that which maximized his expected utility. The individual's expected utility is a function of his income and effort, which in turn are a function of the time-wage and piece rate:

$$W = EU(C, e^2), \quad (12)$$

where $U_1 > 0$, $U_{11} \leq 0$, i.e., the individual is assumed to be risk averse and $U_2 > 0$, $U_{22} \geq 0$. If, after undertaking the job, but before committing himself to his "effort supply," he finds out the value of θ , an individual with a piece rate r and time rate I will choose e so that

$$U_1 r\theta - U_2 = 0. \quad (13)$$

If effort has to be decided before θ is known, we obtain instead¹¹

$$E(U_1 r\theta - U_2) = 0. \quad (13a)$$

Equation (13) (or (13a)) is solved for e as a function of r (r and θ). This is substituted back into (12). Thus, we can write the expected utility as a function of the time rate I and piece rate r :

$$W = EV(r\theta, I), \quad (14)$$

where V is the indirect utility function.¹² The equilibrium contract will be that contract which maximizes (14) subject to the zero profit constraint (11). The first-order conditions yield, after some manipulation, the result that, if e is decided before θ is known,¹³

¹¹ For this subsection, it does not matter which formulation one uses; for nonlinear piece rates the distinction will be important.

¹² In the case of (13a), V is a modified indirect utility function, since e will be independent of θ .

¹³ We maximize

$$EU(e(r, I)\theta r + I, e),$$

subject to the constraint that

$$I \leq (1 - r)e\bar{\theta}.$$

Let λ be the Lagrange multiplier associated with the constraint. Then

$$eEU_C\theta - \lambda(e\bar{\theta} - (1 - r)\bar{\theta}\frac{\partial e}{\partial r}) = 0$$

$$EU_C - \lambda(1 - (1 - r)\bar{\theta}\frac{\partial e}{\partial I}) = 0.$$

$$\frac{1-r}{r} = \frac{-\frac{E(U_C - \bar{U}_C)(\theta - \bar{\theta})}{\theta \bar{U}_C}}{\eta} \approx \frac{R s_\theta^2 r}{\eta}, \quad (15)$$

where s_θ is the coefficient of variation of θ , R is the Arrow-Pratt measure of risk aversion, $-U_{CC}C/U_C$ and η is the compensated supply elasticity of effort, $(\partial \ln e / \partial \ln p)_{\bar{u}}$. Alternatively, letting s_C be the coefficient of variation in consumption,

$$(1-r) = \frac{-\frac{E(U_C - \bar{U}_C)(C - \bar{C})}{\bar{C} \bar{U}_U}}{\eta} \approx \frac{R s_C^2}{\eta}. \quad (16)$$

The piece rate depends simply on the ratio of normalized covariance between θ (or C) and marginal utility of consumption and the supply elasticity of effort. The normalized covariance¹⁴ depends, in turn, on the magnitude of relative risk aversion. Then (15) and (16) give the natural result that the piece rate is higher the smaller the risk, the lower the risk aversion, and the higher the supply elasticity of effort (the greater the incentive effects). Notice that when individuals are risk neutral, or when there is no variance in θ , $r = 1$: individuals simply get paid their marginal product. If relative risk aversion is around unity, the compensated supply elasticity, say 0.5, and the coefficient of variation of consumption 0.4, then approximately a third of income will come from time pay and two thirds from incentive pay. Similar results hold if effort is decided after θ is known.¹⁵

Since

$$dEU = EU_C dI + (EU_C \theta e) dr,$$

$$\frac{\partial e}{\partial r} = \left(\frac{\partial e}{\partial r} \right)_{\bar{u}} + \frac{e EU_C \theta}{EU_C} \frac{\partial e}{\partial I}.$$

Substituting, we obtain (15). To obtain the approximation, we observe that

$$\frac{EU_C(\theta - \bar{\theta})}{EU_C \bar{\theta}} \approx \frac{U_{CC} \bar{\theta} e r}{U_C} \frac{E(\theta - \bar{\theta})^2}{\bar{\theta}^2} = \frac{U_{CC} \bar{C}}{U_C} \frac{r e \bar{\theta}}{e \bar{\theta}} s_\theta^2.$$

¹⁴ The approximation in (15) is precise for a quadratic utility function, where R is taken to be the elasticity of marginal utility at mean income.

¹⁵ We obtain instead of (15)

$$(1-r) = - \frac{\frac{E(\beta - \bar{\beta})(C - \bar{C})}{E \bar{\beta} \bar{E} \bar{C}}}{\bar{\eta}}, \quad (15a)$$

where

$$\beta = \frac{V_I}{\lambda} + (1-r) \frac{\partial e \theta}{\partial I},$$

the marginal utility of income, adjusted for the reduction in firm profits resulting from an increase in income, and

$$E\beta = 1$$

$$\eta = \left(\frac{\partial \ln e}{\partial \ln r} \right)_{\bar{u}},$$

the compensated supply elasticity of effort, and

$$\bar{\eta} = \frac{E e \theta \eta}{E e \bar{\theta}},$$

a weighted average of the compensated supply elasticities. For some simple cases, (15a) may be solved explicitly. Assume there is constant marginal disutil-

□ **Piece rates and self-selection.** When individuals of different abilities are willing to accept a particular contract, there are at least two reasons why a firm has a motivation to hire only the most competent individuals: so long as individuals receive some time rate, i.e., part of their compensation is independent of their performance, then average labor costs will depend on the average productivity of the labor force; and, even if all individuals were paid their marginal products, if there are specific training (hiring/firing) costs, then it is important before those costs are incurred to know the qualities of the individuals being hired. There has recently developed a large literature on the theory of screening,¹⁶ the processes by which individuals are sorted out according to their abilities. Most of this literature has focused on the use of examinations and the use of surrogates for ability, such as education. The object of this section is to show the part that a piece rate system can play in screening, and to show that there is some presumption that the presence of individuals of lower ability leads to a higher piece rate.

Consider the simplest possible case, where there are two ability groups in the population. Individuals know their own ability. In Figure 2a we have plotted their reservation-wage curve, as a function of the time rate and piece rate. Note that as we have drawn the curve, the lower ability individuals have a lower reservation wage if all compensation were paid in the form of time pay (this implies that the opportunity cost of the individuals who are more able at this job is greater, i.e., ability to perform this job is correlated with ability to perform other jobs), but a higher reservation wage if all compensation were in terms of piece rates (implying a comparative disadvantage in the performance of this particular task). Other possible configurations are given by Figure 2b where all individuals have the same opportunity cost of time, and by Figure 2c where individuals who are less able may still have a comparative advantage in the performance of the given task.

FIGURE 2a
RESERVATION WAGE CURVES

FIGURE 2b

FIGURE 2c

ity of labor, constant elasticity of marginal utility of income, and θ has a lognormal distribution, with coefficient of variation s_θ^2 . Then

$$(1 - r) = \frac{1 - (1 + s_\theta^2)^{-\frac{1}{R}}}{\bar{\eta}}$$

$$= \frac{s_\theta^2}{(1 + s_\theta^2) \bar{\eta}}$$

when $R = 1$ (logarithmic utility function). If the coefficient of variation of θ were 0.4, and $\bar{\eta} = \frac{1}{2}$, then the individual would, on average, receive approximately 70 percent of his compensation in the form of incentive pay, and 30 percent in the form of time pay (slightly less than when e was decided prior to knowing θ).

Provided that the response of effort to an increase in the time rate is not too highly correlated with θ , we can approximate (15a) to obtain an expression identical to (16) or

$$\frac{1 - r}{r^2} \approx \frac{R s_Y^2}{\bar{\eta}},$$

where s_Y is the coefficient of variation for productivity ($e\theta$).

¹⁶ See, for instance, Stiglitz [28, 33], Rothschild and Stiglitz [20], Spence [23, 24], Arrow [4], Salop and Salop [21], Akerlof [1].

The expected profits of the firm from hiring a worker of type j are given by¹⁷

$$Ee^j\theta^j - rEe^j\theta^j - I.$$

FIGURE 3
EQUILIBRIUM WITH SELF
SELECTION

Isoprofit curves are then drawn in Figure 3. If there were only the more productive individuals in the economy, the equilibrium would be at the point M_1 , which we have characterized in the previous section; we have marked the corresponding point in Figure 2a. A similar analysis obtains if there were only the less productive workers. In Figure 2a the point M_2 is the point on the low ability reservation curve which maximizes expected profits.

The firm now faces the choice of what contract to offer. It can either attempt to offer a contract which will be attractive to both groups, or it can offer a contract which will attract only one group. Consider, for instance, Figure 2a. M_3 is the intersection between the two reservation wage curves. Any contract on the curve M_3C is accepted only by the most able, and contracts on the curve BM_3 only by the least able. Any contract on the curve AM_3D is accepted by both groups. The optimal contracts if the firm wishes to hire only one quality of laborers is clear: it either offers M_2 , obtaining only low quality labor, or M_3 (a contract on CM_3 involving just a slightly higher piece rate than M_3), obtaining only high quality laborers. Thus, the attempt to hire a higher quality labor force has led to a higher piece rate.

But it is also possible that the firm may find itself in a situation where it cannot (as in Figure 2c) or does not wish to exclude the less able. Then the profitability at any point along the reservation curve of the more able is a weighted average of the profitabilities of the two groups, with the weights being the relative sizes of the two groups (in the applicant pool).¹⁸ By increasing the piece rate, although profitability on the more able individuals is reduced, that on the less able individuals is increased (assuming the supply elasticities do not differ much): again a presumption for a higher piece rate. This may be seen in Figure 3b. Because the isoprofit curve for the less able is steeper than that for the more able,¹⁹ a slight increase in the piece rate

¹⁷ θ^j can be thought of as a summary random variable, representing the effect of all other random variables on the output per unit effort of the j th individual.

¹⁸ In the analysis of Rothschild and Stiglitz [20], there could not, in general, be an equilibrium in which both the more able and the less able work together, i.e., if we apply their analysis of the insurance market to the problem at hand, the only equilibrium would be at point M_3 , where prices of commodities adjust so that the profitability of that contract is zero. The analysis here can, however, be thought of in a somewhat broader context, e.g., where there is not necessarily free entry, so profits may not be zero, and no two firms are quite identical. Thus, each firm takes the reservation wage curves as given.

¹⁹ The isoprofit schedule is defined by

$$\bar{\pi} \equiv (1 - r)Q - I.$$

where Q is his mean output. Thus

$$-\left(\frac{\partial r}{\partial I}\right)\bar{\pi} = \frac{\frac{1}{Q} - \frac{(1-r)}{I} \frac{\partial \ln Q}{\partial \ln I}}{\left(1 + \frac{\partial \ln Q}{\partial \ln r} \frac{(1-r)}{r}\right)}.$$

above M_1 has a negligible effect on the profitability obtained from the more able workers but a strictly positive effect on the profitability obtained from the less able workers. Note that since the less able are above their reservation wage, it is only the more able's reservation wage curve that matters.

More generally, for each piece rate and time rate the firm announces, there will be a different set of applicants. The profitability of the firm can again be written

$$Q(1 - r) - I,$$

where Q , mean output per worker, is just $Ee\theta$, where now the expectation is taken not only over θ for each type applying, but also over the types applying. Thus, an interior solution will entail²⁰

$$(1 - r)\frac{\partial Q}{\partial r} - Q = 0$$

$$(1 - r)\frac{\partial Q}{\partial I} - 1 = 0.$$

Changes in the mean values of output, Q , will be related then not only to incentive effects (r), income effects (I), but also to self-selection effects, i.e., a change in the composition of applicants. In general, this would be expected to increase the effect of an increase in the piece rate on mean output (per worker); an increase in I , however, may now actually increase Q (rather than reduce it); that is, if more able individuals have higher opportunity costs of their time, an increase in I increases the mean quality of the applicant labor force. It is this which enables there to be an interior solution, i.e., one where further reductions in I (for given r) do not eliminate all applicants altogether. This corresponds to the well-known phenomenon that firms pay higher wages than they "have to" in order to obtain a higher quality labor force. This in turn implies that equilibrium may not entail supply equaling demand: the conventional argument that when the supply of labor exceeds demand wages fall to equilibrate the market is not applicable here.²¹

Note, however, that if individuals are perfectly well informed about their own abilities, and there are no other sources of risk, then equilibrium will entail the firm setting $r = 1$, and auctioning off the jobs to the highest bidder, i.e., a *nonpositive time rate*. (If the time rate were positive, then the firm would be making losses.) This result is independent of the distribution of

Thus, provided $\partial \ln Q / \partial \ln I$ and $-\partial \ln Q / \partial \ln r$ are not too different, the smaller Q —the less the productivity—the steeper the isoprofitability curve. Intuitively, an increase in the time rate increases cost per worker the same, regardless of ability, but increases cost per unit of output more the smaller the productivity of a worker. Thus, the reduction in piece rate required to keep the profits the same is larger.

²⁰ This is again an analysis based on a partial equilibrium approach. If there are large numbers of firms which are identical, there may again be a problem with equilibrium, but not only because of the reasons alluded to in note 18 above. Even if the wage itself were the only instrument available to the firm for affecting the quality of its labor force, there might not exist an equilibrium (the Rothschild-Stiglitz analysis assumed that there were two or more instruments available, the price and the amount of coverage). See Stiglitz [31].

²¹ See Akerlof [1] and Stiglitz [33].

abilities or tastes for work in the population; the individual who bids the highest may not be the one who is the most able; the bids will depend on the ability to perform the task relative to his opportunity cost; if all individuals have the same opportunity cost, then it is the most able who obtain the job.²²

□ **Variable tasks with linear incentive schemes.** The jobs considered in the previous two sections could be considered “single task jobs”: the individual takes a job, the job has associated with it a single task, and he does not know how much effort will be required on his part to perform the task.

Many jobs, however, entail the performance of multiple tasks. For instance, an individual in a machine shop will be required to make a variety of kinds of machines. The piece rates may differ, but not in proportion to the effort required by him to perform each task. If there are a large number of tasks, and he can borrow and lend, then the marginal utility of income accruing from his performance of each task will be approximately the same (i.e., he can smooth his income stream), but there will be variability in effort required. Thus, “effort risk aversion” is crucial in determining the equilibrium wage contract, not income risk aversion. Moreover, if the distribution of tasks is given, then by altering the speed with which different tasks are performed, the distribution of time spent performing tasks of different difficulties will be altered. This leads to the result that *there will be a negative time rate*. (Of course, a negative time rate is not usually directly observed; however, by setting a low “base rate” and a high piece rate, if the piece rate schedule were extrapolated back to the “origin” it would imply a negative time rate, as in Figure 4.)

FIGURE 4
IMPLICIT NEGATIVE TIME RATE

The model is a slight modification of that of Section 1. We assume all individuals are identical. θ_i represents the output produced by a unit of effort applied for a unit of time on the i th task; thus $1/\theta_i$ is the time required to complete the task when the individual applies an effort level of e . The individual works for a large firm, and takes the distribution of tasks assigned to him as given, independent of his own level of effort. As we argued earlier, it is reasonable in this context to take a utility function of the form

$$C - Z(e),$$

where $Z(e)$ is the disutility of effort. Then, when assigned the i th task, with a time rate of I and a piece rate of r (we normalized our units so all tasks have the same piece rate), he maximizes

$$I + re\theta - V(e), \quad (17)$$

where we have dropped the subscript on θ . It is convenient to introduce the variable,

²² The model presented in Mirrlees' paper where individuals differ in their abilities but have the same opportunity cost of their time is essentially this special case; thus his result that all individuals get paid their marginal product is equivalent to the result that all except one group are excluded, and this may be done by a linear incentive scheme of the kind discussed here.

$$x = e\theta, \quad (18)$$

the *speed* of production. Then the individual maximizes

$$I + rx - V\left(\frac{x}{\theta}\right), \quad (19)$$

i.e.,²³

$$r = \frac{V'}{\theta}. \quad (20)$$

The firm chooses I and r to minimize its expected costs; since to produce one unit takes θ/e periods, expected costs are just

$$IE \frac{1}{x} + r. \quad (21)$$

It minimizes (21) subject to the constraint that the expected utility per unit of time be equal to its opportunity cost, $\bar{W}$:

$$I + rEx - EV(x\theta) = \bar{W}. \quad (22)$$

Substituting (22) into (21), and differentiating (making use of (20)), we obtain

$$1 - ExE \frac{1}{x} - IE \frac{1}{x^2} \frac{dx}{dr} = 0$$

or

$$I \equiv - \frac{ExE \frac{1}{x} - 1}{E \frac{1}{x^2} \frac{dx}{dr}}. \quad (23)$$

Since $1/x$ is a convex function

$$E \frac{1}{x} < \frac{1}{Ex} \quad (24)$$

so the numerator of (23) is positive; from (20)

$$\frac{dx}{dr} = \frac{\theta^2}{V''} > 0. \quad (25)$$

Hence, the optimal incentive scheme entails $I < 0$. To obtain some order of magnitude of the terms in (23), we take a Taylor series approximation, assume $V(e) = ke^\alpha$, and that θ is lognormally distributed, with coefficient of variation s_θ to obtain

$$\frac{I}{Ex} = \frac{-\alpha \left[(s_\theta^2 + 1) \left(\frac{1+\alpha}{\alpha} \right)^2 - 1 \right]}{(s_\theta^2 + 1) \left(\frac{1+\alpha}{\alpha} \right)^2}$$

The greater the variance of θ and the greater the elasticity of marginal disutility of effort, the more negative is I (relative to total pay, provided $\alpha \geq 1$).

²³ This assumes on each task he knows θ before deciding on e (or he quickly learns θ).

□ **Revision of piece rate schedules.** One of the classical complaints of workers against the piece rate system is that if they work harder, the piece rate is revised. The question is, how can we explain this revision procedure? Surely, if the value of the output of the individual can easily be ascertained, then he should simply be paid according to his marginal product. There are two circumstances in which the value of output of the individual cannot be ascertained:

(1) The individual is working in a group and only the group output is identifiable.²⁴

(2) The individual's physical output can be identified, but there are no prices to guide the firm in determining the value of his output. This is typically the case in very specialized production: there is a market price for the final output, but no price for the output of each stage or component in the production process.

In such circumstances the worker is likely to have better knowledge of the true cost of production (say, in units of labor effort) than the manager. This again can be viewed as essentially a screening problem: the firm attempts to obtain information about the true labor requirements from the individual, to screen the different jobs according to difficulty. The individual has an obvious incentive for informing the manager when the job is difficult (when the piece rate is set too low) but he has no corresponding incentive to inform the manager when the job is easy, i.e., when the piece rate set is too high. Thus, the direct statements of the workers concerning the difficulty of the job which has been assigned to him lack credibility. The firm seeks to infer information about the difficulty of the task from the revealed behavior of the individual; when it obtains this information, it uses it to revise the piece rate.

In the absence of information, jobs of differing difficulties receive the same piece rate; there will be a distortion: the marginal rates of substitution between consumption (income) and effort will differ on different tasks. Although the information reduces this distortion, the method basing the inference about the job difficulty on previous performance has a cost: it reduces the incentive to work; in effect, it lowers the piece rate. Again, the optimal contract will be designed to balance at the margin the benefits and costs of obtaining the information.

The equilibrium contract will be that contract which maximizes expected profit at the exogenously given level of expected utility of the workers. What we wish to show is that, under a simple linear revision scheme, some revision of the piece rate will in general be desirable.

Thus, we consider the model of the previous section, but assume the tasks come in "batches," e.g., the individual is required to produce ten machines of a given type. The firm can announce then that after a given fraction of the machines has been completed, or after a given period of time has elapsed, then it will revise its piece rate, according to a rule based on previous performance.

²⁴ This is the subject discussed in Rothschild-Stiglitz [20] and Akerlof [1].

We need to distinguish two cases: in one, a task is assigned to only one individual. There is no way by comparing the performance of this individual and that of others that the firm can ascertain the true difficulty of the job. The revision of the piece rate *must then depend only on the individual's own output*. In the second case, there is a large group of individuals performing the same task. The revision will depend on the mean output of the group; the individual's income will then depend on the piece rate at his own output. We assume in that case that the group is sufficiently large that the individual ignores the effect of his output on the revision. Analytically, the second case is somewhat easier to handle, and so we discuss it first.

Linear revision schedules with large groups

The analysis for this case is simple because, from the individual's viewpoint, the two periods (before and after revision) are completely separated.

We assume that in the first period the individual's income is given by

$$I_1 + r_1 x_1,$$

where, it will be recalled, x_1 is the speed of production, i.e.,

$$x_1 = e_1 \theta.$$

In the second period the individual's income is

$$I_2 + r_2 x_2,$$

where r_2 is assumed to be a linear function of performance first period (x_1)

$$r_2 = \beta - \gamma x_1. \quad (26)$$

Thus, the i th period the individual maximizes

$$I_i + r_i x_i - V\left(\frac{x_i}{\theta}\right), \quad (27)$$

i.e.,

$$\theta r_i = V'\left(\frac{x_i}{\theta}\right). \quad (28)$$

The firm seeks to minimize expected costs, C . Let N be the total number of units of each task to be performed. Then, $(N - x_1)$ remain to be done after the revision occurs. If the worker works at speed x_2 , it will take him

$$l = \frac{N - x_1}{x_2} \quad (29)$$

time to complete the task. Thus, expected costs can be written

$$EC = E[I_1 + r_1 x_1 + I_2 l + r_2(N - x_1)]. \quad (30)$$

The firm minimizes (29) subject to paying what is required to obtain labor, i.e., expected utility (per unit time) must be equal to the opportunity cost:

$$E \frac{C - V\left(\frac{x_1}{\theta}\right) - V\left(\frac{x_2}{\theta}\right)l}{1 + l} = \bar{W}. \quad (31)$$

Detailed calculations²⁵ show that if we make the first period sufficiently long,

$$\frac{dEC}{d\gamma_{\gamma=0}} < 0, \quad (32)$$

i.e., it always pays to have some revision.²⁶

One of the interesting implications of revision schedules is that, unlike the cases analyzed in Section 1, where it would pay the workers to collude to work harder, here, the workers all believe that they can be made better off if they collude not to work so hard; that is, if they take the payment parameters as given, they observe that by colluding, they can increase their expected utility. They believe that the workers who are working very hard are exerting a negative externality on the rest. But if they do collude, and if there is an elastic supply of labor, the firm will respond to the collusion by changing the parameters of the payment schedule. The firm will observe that the supply of effort functions will not be given by the solution to (28) but by the solution to (34) and (35) below, i.e., taking into account the effect of effort on the revision. The consequence of collusion is that, in the new equilibrium, the workers have the same level of expected utility, but that expected costs of producing have risen. Once the firm has adjusted the payments schedule to take

²⁵ With no revision, $\gamma = 0$, $I_1 = I_2$, $r_1 = \beta$. Hence

$$\left(\frac{dEC}{d\gamma}\right)_{\gamma=0} = \frac{dI_1}{d\gamma} - E \frac{I_2}{x_2^2} (N - x_1) \frac{dx_2}{d\gamma} - Ex_1(N - x_1).$$

$dx_2/d\gamma < 0$ follows directly from (28). Our earlier analysis showed that $I_i \leq 0$, when $\gamma = 0$. Finally, rewriting (31),

$$\Phi \equiv E \frac{I_1 + I_2 l}{1 + l} + \frac{(r_1 x_1 - V(x_1/\theta))}{1 + l} + \frac{l[\beta x_2 - y x_1 x_2 - V(x_2/\theta)]}{1 + l}.$$

At $\gamma = 0$,

$$\Phi_\gamma = E - \frac{x_1 x_2 l}{1 + l} = -E \frac{x_1(N - x_1)}{1 + l}$$

$$\Phi_{I_1} = E \frac{1}{1 + l}.$$

Hence

$$\left(\frac{\partial I_1}{\partial \gamma}\right) = \frac{E \frac{x_1(N - x_1)}{1 + l}}{E \frac{1}{1 + l}} < Ex_1(N - x_1),$$

if the first period is sufficiently long that

$$\frac{dx_1(N - x_1)}{d\theta} = (N - 2x_1) \frac{dx_1}{d\theta} < 0,$$

since $d[1/(1 + l)]/d\theta > 0$, i.e., if the initial period is sufficiently long so that at least half of the task is accomplished before revision.

In the above calculations, the expectation is taken over those values of θ sufficiently small that the task is not completed in the first period.

²⁶ That is, we can find a length of the first period such that for some values of θ revision is required.

account of the collusion, they will all observe that if they fail to collude, expected utility will be below their opportunity cost.²⁷

Linear revision schedules with individualized jobs

Here we consider the case where the jobs assigned to each individual are slightly different. The consequence of this is that individuals are aware of the effect of their effort supply the first period on the payments the second period, and on the length of the second period.

Thus, the individual chooses x_1 and l to maximize

$W =$

$$\frac{I_1 + I_2 l + r_1 x_1 + \beta(N - x_1) - \gamma x_1(N - x_1) - V\left(\frac{x_1}{\theta}\right) - V\left(\frac{x_2}{\theta}\right)l}{1 + l}, \quad (33)$$

i.e.,

$$r_1 - \beta - \gamma(N - 2x_1) - \frac{V'\left(\frac{x_1}{\theta}\right)}{\theta} + \frac{V'\left(\frac{x_2}{\theta}\right)}{\theta} = 0 \quad (34)$$

and

$$I_2 - W + \frac{V'(N - x_1)}{\theta l} - V = I_2 - V\left(\frac{x_2}{\theta}\right) + V'\left(\frac{x_2}{\theta}\right)\frac{x_2}{\theta} - W = 0. \quad (35)$$

Again the firm wishes to minimize expected costs subject to the constraint (31), where now, however, x_1 and l (or x_2) are given by (34) and (35) rather than (28). Again it can be shown that

$$\frac{\partial EC}{\partial \gamma_{\gamma=0}} < 0,$$

if the first period is sufficiently long.²⁸

²⁷ We assume that there is tacit or explicit collusion within the shop, but that there is still a competitive labor market; present workers cannot restrict the supply of labor to the firm.

²⁸ The only alteration in the earlier calculations follows from (35), which we can solve for x_2/θ independent of γ . Thus $\partial x_2/\partial \gamma = 0$. At $\gamma = 0$, $r_1 = \beta$,

$$\frac{\partial x_1}{\partial \gamma} = - \frac{(N - 2x_1)}{\frac{V''}{\theta^2}},$$

$$\frac{\partial x_1}{\partial \theta} > 0, \quad \frac{\partial x_2}{\partial \theta} > 0, \quad \frac{\partial l}{\partial \theta} < 0$$

$$\frac{dEC}{d\gamma} = E \left[I_2 \frac{\partial l}{\partial \gamma} - x_1(N - x_1) + \frac{\partial I_1}{\partial \gamma} \right]$$

$$\frac{\partial I_1}{\partial \gamma} = \frac{E \frac{x_1(N - x_1)}{1 + l}}{E \frac{1}{1 + l}}.$$

The rest of the calculations follow as before.

3. Towards a theory of hierarchy

■ Some recent attempts to explain the hierarchical production structure have been based implicitly on a “conspiratorial view” of economics, that, somehow, the bosses developed a production structure which necessitated the presence of bosses. We do not wish here to evaluate the validity of that view, or to enquire whether a consistent “endogenous technical change” model with such an outcome can be formulated, although these are questions which are perhaps worth pursuing. Rather, I wish to suggest an alternative model in which the hierarchical production relations result from the natural forces of competition.

We need to distinguish between two senses in which the term “hierarchy” is used: (a) vertical hierarchies, i.e., the use of authority (direct commands) as opposed to monetary or other forms of incentives; and (b) horizontal hierarchies, i.e., the use of differential treatment of individuals, in terms of status, pay, etc. Here we are concerned only with the former.

Vertical hierarchies necessarily entail the existence of supervisors. In Section 5, the supervisor is viewed as obtaining information that formerly was lacking; whether it is economically efficient to have the supervisor depends on the value of this information relative to the costs of the supervisor.

Another set of functions of supervisors relates to his role as a decision maker; Section 5 also develops an argument for some centralization of decisionmaking on the basis of certain nonconvexities associated with information.

□ **Supervisors and imperfect information.** We began Section 2 with a list of the primary determinants of the choice between piece and time rates. Two of the reasons for using the piece rate system relate to a lack of information, which a supervisor could at least partially remedy: (a) incentives are required because inputs cannot be directly observed (supervisors as monitors of inputs); (b) the piece rate system affects the quality of the labor force, when the quality of individuals cannot be directly observed (supervisors as monitors of ability). Two of the limitations on the use of the piece rate system may, at the same time, be alleviated by supervisors: (a) supervisors may serve as monitors of output, e.g., ensuring that quality standards are adhered to, or, in situations where only the group output is easily observable,²⁹ ascertaining the contribution of different members of the group; (b) supervisors are required to determine the appropriate piece rates. The supervisors are thus acting as a substitute for the price system (which would provide the appropriate piece rates).

In the remainder of this section, we elaborate on the role of supervisors as monitors of inputs. Section 2 established that if the effort of the individual could not be directly observed, then the equilibrium contract would entail a piece rate in addition to a time rate. The outputs were random, but because inputs could not be identified, the individual could not obtain complete insurance for his risky income. At the same time, the partial insur-

²⁹ These kinds of production activities have been extensively discussed by Akerlof [1] and Stiglitz [33] in the context of self-selection mechanisms for identifying individuals' abilities.

ance that he did obtain reduced his work incentives. These are the two costs of the incomplete information: the greater risk absorbed by the individual and the lower work effort. Information about inputs may be obtained—but at a cost: the cost of the supervisor who monitors the inputs of the individual.

There are several aspects of the individual's performance that the supervisor might monitor. When output and θ are observable, e can be inferred and if output and e are observable, θ can be inferred. This immediately enables a new range of contracts to be considered.

The simplest would have the individual agree to exert a given amount of effort. The output would then depend on θ and e . His pay would not, however, depend on output. Were such a contract feasible, the individual would (in an *ex ante* sense) be better off than he would be with the contracts discussed earlier. For he can now completely eliminate all risk. The optimal contract is given by

$$\theta U'(C) = V'(e), \quad (42)$$

where

$$C = e\bar{\theta}.$$

If $U''' > 0$, $U'\theta$ is a concave function of θ ,

$$EU'\theta < U'(e\bar{\theta})\bar{\theta}, \quad (43)$$

and the average *income* of individuals will be higher in this contract than in the ones described in Section 1.³⁰

Workers voluntarily undertake to be supervised; a certain amount of compulsion will be a characteristic of competitive equilibrium. They each are working harder than the incentive system itself provides an inducement to work. They submit to being compelled to work harder than direct incentives provide for, because the consequence is a higher expected utility.³¹

³⁰ To assure the concavity of $U'\theta$, we require, if e is decided after θ , that $d^2e/d\theta^2 > 0$.

³¹ A still better contract from an *ex ante* point of view is one in which the individual commits himself to a level of effort as a function of θ . If such a contract were enforceable, the optimal contract would entail the individual's receiving the same income regardless of the θ he faced, but those who experience higher θ (e.g., were assigned a better piece of land or an easier job) work harder. That is, the individual would sign the contract that maximized

$$EU(C(\theta)) - V(e(\theta)), \quad (i)$$

where

$$EC(\theta) = Ee(\theta)\theta. \quad (ii)$$

The solution to (i) entails

$$U' = \text{constant for all } \theta$$

and

$$\frac{V'}{\theta} = \text{constant for all } \theta.$$

Those with higher θ have a high marginal disutility of effort, i.e., they work more. Their *ex post* utilities are actually lower than those with low ability. These results are exactly parallel to Mirrlees' results for the optimal distribution of income in a utilitarian framework in a completely controlled economy. When θ represents the individual's ability to perform a particular job, the perverse incentive effects (i.e., individuals' hiding their abilities) are even stronger with this contract than with that previously discussed.

Although each worker may resent this compulsion and feel it is unnecessary on his own part, he prefers to work for firms which use this compulsion, recognizing that without it, some of his colleagues will slough on the job, and thus firms which employ some degree of compulsion are able to pay higher wages.

This contract appears reasonable when the θ which is being monitored is an exogenously determined risk variable (the weather, the quality of the machine which the individual has to work on, etc.); but where the risk variable is the individual's ability to perform the given task, there would be a strong incentive for the more able individual to pretend that he was not particularly capable and thus economize on his effort. Some levels of incentive schemes would still be required.

In deciding on whether to accept a contract with this kind of compulsion, individuals who dislike authority relationships may require considerable additional compensation to induce them to accept this kind of employment relationship. One of the disadvantages of this kind of relationship is that it may not allow the kind of individual variability (e.g., on a day-to-day basis) that the piece rate system allows. The viability of such contracts is then a reflection of the fact that for some individuals, these disadvantages are outweighed by the greater wages and more certain incomes that firms who use these contracts can pay.

There is some ambiguity about whether the relationships defined in this section ought to be referred to as authority relationships. Incentive structures such as those discussed in Section 2 clearly do not involve authority relationships. On the other hand, extreme nonlinear incentive structures, which take the form: if the worker does not complete this task, he will receive no pay or will be fired, might or might not be thought of as an authority relationship. If there were no uncertainty about whether or not the task was completed, then there is no discretion on the part of the supervisor, and it is unlikely that we would refer to this as an authority relationship. But if there is uncertainty about whether the task is completed, then the supervisor is, in effect, in the position of whether to authorize payment to the individual. Since payment will, as a result, be stochastic, the amount (or quality) of supervision will affect both the optimal incentive scheme which will be used and the level of expected utility which the individual will attain. Thus the analysis presented here and in Section 2 may be viewed as presenting two polar cases—where there is no monitoring of input, and perfect monitoring of input. Most employment contracts involve some, but not complete monitoring.³²

□ **Supervisors and intervention.** The function of supervisors in making decisions, intervening in the actions of the subordinates, giving commands, is perhaps the aspect that is most central to the radical interpretation of the role of supervisors. The previous section delineated this role as a contract enforcer.

There is, however, more to the role of the supervisor as a

³² For a more extended discussion of these issues, see Stiglitz [32] and Mirrlees [16].

giver of commands than simply that of contract enforcer. All agents in the economy possess a certain degree of autonomy in performing their tasks; on the other hand, there are conditions under which there is some *intervention* with the autonomy. It is this aspect of supervisors which I wish to discuss here.

We can imagine a number of ways in which a firm can respond to a new or unusual contingency or to new information:

(1) The firm designs a set of incentive structures, such that as contingencies or new information arises, individuals respond in the way that is optimal from the point of view of the firm, without direct monitoring of the worker.

(2) The firm designs a set of routines (commands), which distinguish among the different contingencies, and a monitoring-reward structure that induces individuals to follow the routines.

(3) The firm instructs the individual to consult the supervisor when a particular set of contingencies or a particular category of new information arises; the supervisor then makes a decision, instructing the worker what to do.

(4) The firm maintains a monitor on the various contingencies (or on new information); when contingencies for which the worker should take an action other than that provided for by the ordinary incentive structure (or the ordinary set of commands) occur, the supervisor intervenes and issues a new set of commands or a new incentive structure.

(5) Finally, the firm can simply fail to take cognizance of the contingency, treating this "state of nature" as if it were the ordinary state of nature.

The first two responses involve the supervisor only in the roles which we discussed earlier;³³ the third involves the supervisor in the role as a consultant; in the fourth the supervisor directly intervenes in the actions of the worker, and it is in this role that the authority relationship is most clearly exercised. Thus, in the fourth organizational form, there is intervention by the supervisor in the autonomy which the individual has in the performance of his tasks. Each of these methods of responding to a particular contingency has its costs and benefits.

What I wish to argue in this section is that these information flows involve a number of different nonconvexities; these nonconvexities give rise in turn to specialization, and it is this which provides the economic justification for the authority relationships. (The choice of organizational form may depend, of course, on a number of other factors, e.g., workers' attitudes towards authority. These aspects of the problem have been extensively discussed in the literature on organization theory.³⁴ The discussion of this section should thus be seen as complementing this literature, in focusing on an aspect of the problem which has perhaps received insufficient attention.)

³³ This routinization is close to what March and Simon call performance programs.

³⁴ See, for instance, March and Simon [12].

Benefits

The return to intervention is that the supervisee, without intervention, may take actions which are not in the interests of the firm. Now in the case of *routinized actions*, those which occur repeatedly, this is no more than another aspect of contract enforcement. But the class of events we are interested in here are those which occur in response to events which occur sufficiently infrequently that the firm has not specified the correct action to be taken by the supervisee, and/or has not designed the incentive scheme to take these events into account appropriately. Many of these events are of relatively little importance: the worker on the assembly line observes a part which is defective in a way which is unusual. He must make a judgment about whether the part should be discarded or used. There may be some incentive for the individual to attempt to make the correct judgment:³⁵ if he does, he is more likely to be promoted. But even if he has no incentive, and simply decides randomly, it may not pay the firm to intervene in the action; for the time required for the supervisor to come to the assembly line, examine the part, and make a decision is simply not worth it. There are, however, other nonroutinized events which are obviously important, and for which intervention would clearly be desirable.

Costs

We need to distinguish between several categories of costs of the various organizational forms: (a) The fixed costs of setting up the system. This is primarily associated with the design of the appropriate routines and incentive structures. (b) The costs of training workers, of teaching them how they should respond in different contingencies. (If workers are good optimizers, then all they have to learn is the complete incentive structure, but when there are a number of different kinds of actions which the worker can take, the incentive structures may be very complicated.) Presumably, it is less costly to teach them to act as an information filter, i.e., to recognize certain classes of contingencies to refer to the supervisor, than to teach them to both recognize the contingencies and take the correct actions. (c) The cost of continuous monitoring, in the case where the supervisor must intervene in the actions of the worker. (d) The cost of having the supervisor available as a consultant. (e) The variable cost which arises when the contingency occurs. Thus, in organization (3), where the supervisor acts as a consultant, there is both a cost of the worker consulting the supervisor, and the supervisor issuing a command to the worker, while in organiza-

³⁵ A central aspect in which the analysis in this context differs from the theory of teams is that the workers and the managers (owners of firms) do not necessarily have the same objectives. There is not only a question of efficient transmission of information, but also—as we noted in our earlier discussion where workers know the difficulty of the job they undertook—of inducing individuals to reveal this information. Our earlier discussion also provided an example of a (second best) method of inducing individuals to reveal information to the firm, which could then use the information to provide a “better” incentive scheme.

tion (4) there is only the cost of communicating the decision from the supervisor to the worker. The first set of costs are once and for all costs. The second set is a cost per unit time, but the magnitude of the costs depends on the rate of labor turnover. The costs (c) and (d) are fixed costs per unit time, but the magnitude of these can vary, depending on the quality of services to be provided by the supervisor. Finally, cost (e) is a variable cost. Thus, to a large extent, the choice of organization forms can be viewed as a tradeoff between fixed and variable costs.

The costs of routinization are likely to be independent of the gains to be had from routinization, so that it only pays to undertake these large fixed costs for events for which the expected gain—i.e., the value of the gain times the expected frequency of the contingency's occurring—are large. On the other hand, the choice of whether to employ the supervisor as a consultant or to have the supervisor monitor the status and intervene depends on the relative size of the fixed costs (training costs in one case, versus monitoring costs in the other) and the expected variable costs (which depend in turn on the probability of occurrence times the difference in the variable costs, when the contingency arises). For events which occur sufficiently infrequently that the expected variable costs are low, intervention would appear optimal. Rather than undertake the large costs of training individuals to recognize contingencies which occur relatively infrequently, it may be preferable to have an individual who specializes in the acquisition and transmittal of information.³⁶

This argument, as plausible as it seems, becomes less convincing when we allow the amount of information to be variable: why not give a little information to a large number of individuals rather than give a lot of information to a few? To have the kind of specialization in economic activities that the above analysis suggests requires a nonconvexity. In fact, there are two nonconvexities associated with information. The first is that the

³⁶ More formally, let B_i be the gross benefit occurring under the i th organization form, R_i be the fixed costs of routinization (for organization forms 1 and 2), t_i be the cost of training individuals under the i th organizational form, S_i/π be the supervisory costs, v be the cost of transmitting a decision from the supervisor to the worker, u be the cost of transmitting information from the worker to the supervisor, N be the number of workers, π be the frequency of the event, and $\rho(N)$ be the probability, given that the event has occurred, that the supervisor will detect it. Thus, to decide what kind of organization to employ, we compare

$$\begin{aligned} \pi B_1 - R_1 - N t_1 - S_1 \pi, \\ \pi B_2 - R_2 - N t_2 - S_2 \pi, \\ \pi B_3 - N t_3 - \pi(u + v) - S_3 \pi, \\ \rho \pi (B_4 - v) + (1 - \rho) \pi B_5 - S_4 \pi, \\ \pi B_5. \end{aligned}$$

Normally, we expect

$$\begin{aligned} R_1 > R_2, \quad S_1 < S_2 < S_3 < S_4, \\ B_5 < B_1 < B_2 \leq B_3 = B_4, \end{aligned}$$

and

$$t_3 < t_1, \text{ and } t_3 < t_2.$$

The dependence of the choice of organizational form on the various parameter follows in a straightforward manner.

value of a given amount of information increases with how frequently it can be used; that is, even if the initial cost of acquisition of information of the employee and the supervisor were the same, in a hierarchical structure the supervisor uses the information more frequently. Unlike conventional commodities, each use does not detract from the “stock” of information he has—and indeed, to the extent that there is learning, it may actually increase it. Although this provides an argument for specialization, the advantages of specialization may be offset by the fact that workers may find it easier to acquire certain categories of information: a naturally joint product of the performance of their tasks is the acquisition of certain information (e.g., observations about defects in the materials with which they are working).³⁷ On the other hand, many categories of information (e.g., concerning market phenomena) can more easily be obtained as joint products of other activities in which the supervisor is engaged; moreover, if the information impinges on the actions of a large number of individuals, it may be less expensive for the supervisor to acquire the information and then transmit it to the workers, than to have each worker acquire it himself.

Secondly, it can be shown that in general, it does not pay to acquire just a little bit of information: a little information has negative net value, i.e., the costs always exceed the benefits.³⁸ Thus it does not pay to monitor just a little; it either pays not to monitor at all, or to monitor at a finite level. This, I suggest, may provide a framework not only for defining the degree of autonomy of individuals within their jobs, but also for defining the boundaries of firms.³⁹

The literature on organization theory⁴⁰ has long recognized the costs of acquiring and disseminating information, the role of supervisors in this process, and the bounds on rationality which result from the fact that information is costly. What we have attempted to do in this section is to relate supervision to the structure of the technology of information, to argue that fundamental nonconvexities associated with information result in the specialization in functions which is implicit in the hierarchical relationship.⁴¹

4. Concluding remarks

■ The problems we have analyzed in this paper can be viewed as special cases of the class of problems which we have discussed under the rubric of “the Theory of Screening.” The essen-

³⁷ Much of the literature on decentralization has focused on this case, where information naturally arises in the periphery, and then flows to the center.

³⁸ See Radner and Stiglitz [19].

³⁹ The characteristic of independent firms is that the only intervention is provided automatically—when the firm is unable to pay its debtors. The merging together of two firms requires the expenditure of a continuing fixed cost thereafter by the parent firm in monitoring the performance of the acquisition.

⁴⁰ See, for example, March and Simon [12], and Williamson [34].

⁴¹ The argument, of course, is only a qualitative one; that is, one can show that under quite general conditions there will exist some nonconvexities. To explain specialization requires that these nonconvexities be in some sense significant.

tial problem of the theory of screening arises from the inability of some agent in the economy (the insurance firm, the government, the employer) to distinguish costlessly among objects which are, or are potentially, different.

Here we have been concerned with the problems associated with distinguishing levels of effort and ability of different individuals and distinguishing the level of difficulties associated with different jobs.

Of course, not all kinds of imperfect information can be cast into the screening framework, and indeed, the discussion of Section 3 was intended to delineate those cases where the imperfect information could and could not be viewed as a screening problem.

Most of the paper was concerned with how firms (and their workers) came to terms with imperfections of information of the kind associated with screening. The theory of screening has identified several different ways in which the economy responds to the absence of this kind of information:

(1) The firm may attempt to reduce the costs imposed by the absence of information. Coinsurance clauses in insurance policies make it less important for the insurance company to observe the actions of the individual; here incentive pay makes it less important to observe the level of effort of the individual.

(2) The firm may attempt to obtain the information indirectly, to infer the information from the actions taken by the individual. Examples of this are given in Section 2, where the performance of the individual reveals information about the difficulty of the job or his ability.

Thus, most of Part 1 could be considered an exploration of these two alternative approaches to the existence of imperfect information. Part 2 on the other hand, was concerned with the alternative approach.

(3) There may be a direct expenditure in observing the differences; this is analogous to "screening by examination" in the context of education. Here we have identified direct examination with the supervisor; he observes how hard the individual is working or the ability of the individual.

This paper has thus been concerned with the demand for supervision and the return to having a hierarchical production structure. We have argued that individuals without information about their abilities to perform a particular job (or who face other risks) would be willing to pay for supervision to ensure that they and others work harder than they have a direct incentive to work; that individuals with information about their abilities⁴² are willing to pay to have supervision to ascertain these abilities so that they can capture their ability rents; and that for individuals who have information about their abilities but face other risks affecting their marginal productivities, there is a return to a hierarchical production structure.

The desirability of the hierarchical production structure

⁴² If they believe their abilities to be above average. Similar results obtain for individuals who are less averse to working hard

must, of course, be related to the value of supervision in relation to its cost, and in this paper we have ignored the costs of supervision. A theory which purports to explain the evolution of the hierarchical structure of production must be concerned with how changes in technology affected the costs and benefits of this method of organizing production. These are questions which will have to be pursued elsewhere.

References

1. AKERLOF, G. "A Theory of Information and Labor Markets." *Quarterly Journal of Economics* (forthcoming, 1975).
2. ALCHIAN, A. AND DEMSETZ, H. "Production, Information Costs, and Economic Organization." *The American Economic Review*, Vol. 62, No. 5 (December 1972).
3. ANAND, S. "An Optimal Earnings Schedule." Oxford, mimeo.
4. ARROW, K. J. "Higher Education as a Filter." *Journal of Public Economics*, Vol. 2 (July 1973).
5. ———. *The Limits of Organization*. New York: W. W. Norton, 1974.
6. ATKINSON, A. AND STIGLITZ, J. "Alternative Approaches to the Redistribution of Income." Stanford mimeo, 1975.
7. CYERT, R. M. AND MARCH, J. G. *A Behavioral Theory of the Firm*. Englewood Cliffs, N.J.: Prentice Hall, 1963.
8. DOERINGER, P. AND PIORE, M. *Internal Labor Markets and Manpower Analysis*. Lexington, Mass.: D. C. Heath, 1971.
9. HARRIS, J. E., WACHTER, M. L., AND WILLIAMSON, O. E. "Understanding the Employment Relation: The Analysis of Idiosyncratic Exchange." *The Bell Journal of Economics*, Vol. 6, No. 1 (Spring 1975).
10. LERNER, A. P. *The Economics of Control*. New York: 1944.
11. LUPTON, T., ED. *Payment Systems*. Penguin Modern Management Readings, 1972.
12. MARCH, J. G. AND SIMON, H. A. *Organizations*. New York: John Wiley, 1958.
13. MARGLIN, S. "What Do Bosses Do?" Harvard University mimeo.
14. MIRRLEES, J. A. "An Exploration in the Theory of Optimum Income Taxation." *Review of Economic Studies*, 1971.
15. ———. "Notes on Welfare Economics, Information and Uncertainty," in Balch, McFadden and Wu, eds., *Essays on Economic Behavior under Uncertainty*, Amsterdam: North Holland, 1974.
16. ———. "The Optimal Structure of Incentives and Authority Within an Organization." Oxford mimeo.
17. PENCARVEL, J. H. "An Essay on the Economics of Work Effort and Wage Payments Systems." Stanford mimeo, July, 1974.
18. PIGOU, A. C. *The Economics of Welfare*. London: MacMillan, 1920, Chapter VIII.
19. RADNER, R. AND STIGLITZ, J. "A Nonconcavity in the Value of Information." Stanford mimeo, 1975.
20. ROTHCHILD, M. AND STIGLITZ, J. E. "Equilibrium in Competitive Insurance Markets; the Economics of Markets with Imperfect Information." *Quarterly Journal of Economics* (forthcoming 1975).
21. SALOP, J. K. AND SALOP, S. C. "Self Selection and Turnover in the Labor Market." Paper presented to Econometric Society Meetings, Toronto, 1972.
22. SAMUELSON, P. A. "A. P. Lerner at Sixty." *Review of Economic Studies*, Vol. 31 (June 1964), pp. 169–178.
23. SPENCE, A. M. "Competitive and Optimal Responses to Signals: An Analysis of Efficiency and Distribution." *Journal of Economic Theory*, Vol. 7 (March 1974).
24. ———. "Job Market Signalling." *Quarterly Journal of Economics*, Vol. 81 (August 1973), pp. 355–379.
25. ———. *Market Signalling*. Cambridge: Harvard Univ. Press, 1974.
26. ——— AND ZECKHAUSER, R. "Insurance, Information, and Individual Ac-

- tion. *The American Economic Review*, Vol. 61, No. 2 (May 1971), pp. 380-387.
27. STIGLITZ, J. E. "An Economic Analysis of Labor Turnover." Stanford Working Paper, 1975.
 28. ———. "Equilibrium Wage Distributions." Cowles Foundation, 1974.
 29. ———. "Information and Economic Analysis." Paper presented to AUTE Conference, Manchester and published in AUTE Proceedings, M. Parkin, ed.
 30. ———. "Labor Turnover and Wage Structure," Yale mimeo.
 31. ———. "A Market with Heterogeneous Qualities." Stanford mimeo.
 32. ———. "Risk Sharing and Incentives in Sharecropping." *Review of Economic Studies* (April 1974).
 33. ———. "The Theory of Screening, Education, and the Distribution of Income." *The American Economic Review* (forthcoming 1975).
 34. WILLIAMSON, O. E. *Corporate Control and Business Behavior*. Englewood Cliffs, N.J.: Prentice Hall, 1970.
 35. WILSON, R. "On the Efficient Scale of a Firm." Stanford University. August 1974.