Optimal Contracts for Teams

R. Preston McAfee; John McMillan

International Economic Review, Volume 32, Issue 3 (Aug., 1991), 561-577.

Stable URL:
http://links.jstor.org/sici?sici=0020-6598%28199108%2932%3A3%3C561%3AOCFT%3E2.0.CO%3B2-1

INTERNATIONAL ECONOMIC REVIEW
Vol. 32, No. 3, August 1991

OPTIMAL CONTRACTS FOR TEAMS*

BY R. PRESTON MCAFEE AND JOHN MCMILLAN[1]

In a team subject to both adverse selection (each member's ability is known only to himself) and moral hazard (effort cannot be observed), optimal contracts are, under certain conditions, linear in the team's output. The outcome is the same whether the principal observes just the total output or each individual's contribution. Thus monitoring is not needed to prevent shirking by team members; instead, the role of monitoring is to discipline the monitor.

1. INTRODUCTION AND SUMMARY

Production is a collective enterprise: by workers in a firm or firms in a joint venture. The synergy that is the reason for the team's existence may mean that an individual's contribution to the team's output is not distinguishable, so that it is impossible to pay him according to his own productivity. How should the principal remunerate the team members so as to maximize his own profit? In this paper we analyze incentives in teams with asymmetric information: both adverse selection (each member's ability is known only to himself) and moral hazard (effort cannot be observed directly).

Holmstrom (1982) showed that, in a team model with moral hazard, the principal can ensure an outcome arbitrarily close to the full-information ideal by using a contract that punishes each team member arbitrarily severely whenever team output falls below some target. However, this seems to be an unrealistically drastic way of solving the moral-hazard problem.[2] In our model, we in effect prevent the principal from using such a contract by introducing adverse selection in addition to moral hazard, and by assuming that ability and effort interact in such a way that the principal is unable to disentangle an agent's effort from his ability. Under the assumption that the principal and the team members are risk neutral, we shall show that the principal can implement his information-constrained optimum simply by offering each agent a payment linear in the team's output.

Although the essential feature of our analysis is the interaction between adverse selection and moral hazard, this result can be understood by first considering the special case in which the agents' abilities are common knowledge. In this case the moral-hazard problem can be completely solved: the principal can do as well as he

* Manuscript received December 1989.

[1] We thank Hideshi Itoh, Paul Milgrom, John Riley, and two referees for comments. McMillan thanks the National Science Foundation (SES 872 1124) for research support.

[2] Also, as Arrow (1985) pointed out, this solution suffers from a multiple-equilibrium problem. Many combinations of agents' actions are best responses to this contract: "If some individuals shirk a little, it pays the others to work somewhat harder to achieve the same output. Hence the scheme does not enforce the optimal outcome, though it permits it" (p. 47).

could if he were in a full-information world and paying each agent his marginal product. The principal offers to pay each of the n agents 100 percent of any marginal increase in team output. Clearly this gives each agent the appropriate incentive to exert effort. It does, however, result in the principal's total variable payment being n times the value of output. To balance this, the fixed part of the payment function must be negative: in fact, in this case each agent's fixed payment is set equal to the expected value of output minus the agent's production cost, so that the agents earn zero rents on average. Thus the optimal contract has the principal initially (before production takes place) collecting money from the agents, and then (after the production process) paying each agent the full value of the team's output.[3] In other words, in the presence of moral hazard, the principal achieves his ideal outcome by, at the margin, breaking the budget: by eliminating the requirement that the marginal payments sum to one, and by manipulating the lump-sum payments instead of the marginal payments (as Groves 1973 first noted).

Now add adverse selection to the model. The privacy of the agents' information about their abilities results in their earning informational rents. The principal extracts some (but not all) of these informational rents by reducing the marginal payments below 100 percent: thus he distorts the outcome away from the first-best. As the variable payments are reduced, the fixed payments become less negative; that is, the agents' initial payments are smaller. If the uncertainty about the agents' abilities were sufficiently dispersed, the marginal payments would become so small that they would sum to less than one, and the fixed payments would become positive: this now looks like a more conventional payment scheme of salary plus commissions (Theorem 2 in Section 2).

Why does the principal deliberately reduce the total gains from trade by paying less than full marginal value product? The answer comes from the privacy of the agents' knowledge about their own abilities. There is a trade-off between the adverse selection and the moral hazard. If the principal were to pay full marginal value product to all agents, he would not be able to distinguish between high-ability agents and low-ability agents. He would have to offer a sufficiently generous contract to induce low-ability agents to participate; but this would leave high-ability agents with large rents. By reducing marginal payments below 100 percent, and offering different margins to different agents, the principal can sort the agents by their ability. At the cost of lowering the total gains from trade, he reduces the rents earned by more able agents. By appropriate choice of contracts to offer, the principal can ensure that the extra rents he extracts from the able agents exceed the reduction in total rents; the distortion from paying less than full marginal value product is optimal from the principal's point of view.[4]

The principal implements his information-constrained optimum by asking the agents to reveal their abilities. The marginal rate of payment offered to any one agent depends not only on his own reported ability but also on the reported ability levels of all the other agents. Thus an agent is induced to exert more effort (a) the

[3] Holmstrom (1982, p. 328) briefly discussed a contract of this sort for the case of no uncertainty.

[4] For a more detailed account of the inefficiency arising from the trade-off between adverse selection and moral hazard, see McAfee and McMillan (1987b, pp. 299–300).

greater his own ability; and (b) the greater the ability of any of his team-mates. An agent's rents increase both with his own ability and with the other agents' abilities. By structuring the payment scheme in this way, the principal ensures that the agents find it in their interest to report their abilities honestly (see Theorems 2 and 3 in Section 2). Thus the adverse selection leaves the agents earning positive rents. The interdependencies among the agents' payment functions mean that an agent cares about his team-mates' abilities. The more able are his team-mates, the harder an agent will be induced to work; but this is more than compensated for by his increased payments.

Unless the team-members' abilities are potentially widely spread, the optimal contract based on team output has the property that, at the margin, the principal pays out more in bonus payments than the marginal value of output. (This is because the principal wants to pay each individual close to his own marginal value product.)

Suppose now that the principal can costlessly monitor the individuals' contributions to team output. Our analysis yields the surprising result that the principal can do no better when he monitors than when he simply bases payments on team output as described (Theorem 4 in Section 3). This is most easily seen in the special case in which adverse selection is absent: there, as we saw, the first-best outcome can be attained using payments based on total output, and obviously monitoring cannot improve upon this. Theorem 4 shows, less trivially, that the same result holds when information about abilities is private. In other words, our model contradicts the conventional wisdom that piece rates, where feasible, are inherently more effective than group-payment schemes. In our model, an appropriately designed group scheme will work as well as a piece-rate scheme.

Hence the jointness of production and the unobservability of individuals' contributions need not create a free-rider problem. The principal can do as well when he observes only total output as when he observes the individual contributions. Monitoring is not needed to prevent shirking by the team members. Thus our results agree with Holmstrom (1982) in contradicting the assertions of Alchian and Demsetz (1972) that "each input owner will have more incentive to shirk when he works as part of a team, than if his performance could be monitored easily or if he did not work as a team" and that the purpose of monitoring is "disciplining the team" (p. 780).[5]

What then is the purpose of monitoring? Monitoring affects the principal's incentives. Paradoxically, the role of monitoring is to discipline the monitor. Recall that the optimal contract based on team output often (but not always) has each team member making an initial payment to the principal and then the principal paying the agents a sum greater than the value of the output. With monitoring, and payments based on individual contributions, the net payments are the same as without monitoring, but the gross payments are smaller. We shall find that, with monitoring, the principal pays ex post a sum less than the total value of output, and therefore keeps part of the output for himself: the principal has a positive stake in the outcome (Theorem 6 in Section 3). When the agents make ex ante payments to the

[5] A similar result has been obtained independently by Picard and Rey (1987).

principal and the contractual ex post payments sum to more than the value of output, the principal is subject to moral hazard: it is in the principal's interest to renege on his contract by somehow sabotaging the production process.[6] The principal is tempted to take the money and run. By the foregoing argument, such a situation does not arise when there is monitoring, although it often does under optimal contracts in the absence of monitoring. Our analysis assumes away any possibility of the principal's reneging by assuming that the principal is able to commit himself to the terms of the contract. Suppose, however, we were to make the commitment endogenous. A common way of achieving commitment is through the reputational effects inherent in repeated plays of the game. The principal does not renege because he knows that reneging will destroy his reputation and thereby reduce his bargaining power for the future. (''Reputation, i.e. credibility, is an asset''—Alchian and Demsetz 1972, p. 778.) The larger the immediate gains from reneging, the less likely they are to be outweighed by the future losses. Thus commitment is more credible the smaller the gains from reneging. It follows from our analysis that the principal is more likely to maintain his commitment when he monitors than when he does not. In this sense, the purpose of monitoring is to discipline the monitor.[7]

2. OPTIMAL CONTRACTS BASED ON TEAM OUTPUT

The team consists of a principal and n agents. All are risk neutral. Each agent is endowed with an ability level z_i, $i = 1, \dots, n$. Only agent i himself knows z_i; the other agents and the principal perceive z_i to be drawn independently from a distribution $G(z_i)$. Assume that the density $g(z_i) = G'(z_i)$ exists and is continuous. Choose units for z such that the support of G is $[0,1]$. Note that we have ruled out by assumption correlations among the agents' abilities. Each agent chooses a level of effort (possibly vector-valued) which is not directly observable by anyone else. Output x, which is observed by the principal, depends upon the team members' efforts and abilities, as well as a random variable. We leave implicit in the notation both the efforts and the random shock. We suppose that agent i's effort and ability combine to produce his contribution to output measured in efficiency units, y_i. The monetary cost to the agent of this effort is $c(y_i, z_i)$, with $c_y \geq 0$, $c_{yy} \geq 0$, $c_z < 0$, and $c_{yz} \leq 0$ (where subscripts denote partial derivatives). That is, both cost and marginal cost decline as ability rises. (Thus z should be interpreted as an ability specific to this production process, for with general abilities, opportunity-cost considerations would tend to make c_z positive.) Let $y = (y_1, \dots, y_n)$ denote the vector of agents' contributions and suppose that the density of output x given y can be written as $f(x, y)$. This formulation embodies the assumption that the principal, observing either output x or (as in the next section) individual contributions y_i, is unable to disentangle effort from ability. Ability and effort affect output

[6] As Eswaren and Kotwal (1984) pointed out, the principal also faces moral hazard in operating the contract with group penalties of Holmstrom (1982).

[7] Note that with monitoring, some commitment is still needed: the principal still has a short-run incentive after production has taken place to fail to pay the agents as promised. But the point is that, with monitoring, the temptation to renege is smaller and so the commitment is more likely to be maintained.

jointly in the form of efficiency units. This assumption, which will be crucial to our results, has been used before by Laffont and Tirole (1986, 1987) and McAfee and McMillan (1987b).[8]

We confine attention to mechanisms in which the agents directly report their types to the principal, since by the Revelation Principle (see Myerson 1982, for example), any mechanism can be mimicked by such a direct mechanism. Let $\hat{z}_i$ denote i's report to the principal about his ability z_i.

The principal commits himself in advance to the rule by which he will pay each agent. Payment can be a function of what the principal observes, which is the final output x and the reports $\hat{z}$; denote the payment to agent i by $p_i(x, \hat{z})$. Note that the payment to one individual can, and in the optimum contract will, depend on all the others' reports.

The principal's payoff is the total value of the team's output minus his payments to the agents:

$$\phi = U(x, y, z) - \sum_{i=1}^{n} p_i(x, \hat{z}). \tag{1}$$

This formulation allows the possibility that the principal cares not only about the aggregate output but also the individual inputs of effort and ability. For example, the agents' efforts or abilities may affect not only the quantity of the final output, x, but also its quality; but it may be impossible to make payment contractually dependent on quality. Also, the principal's payoff may be nonlinear in output x: for instance, the principal may be a monopolist in the product market, so that $U(x) = xD(x)$ where $D(x)$ is the demand curve for the output.

The agent's payoff is his payment less his cost of effort:

$$\pi_i = p_i(x, \hat{z}) - c(y_i, z_i), \qquad i = 1, \dots, n. \tag{2}$$

Denote by $\mu(y)$ the expected value of x. Assume $\mu_i > 0$ (where the subscript denotes the partial derivative with respect to y_i). Assume $\mu_{ij} > 0$ for $i \neq j$; that is, individuals' contributions are complements in production.[9] Also, assume μ is concave in y, so that there are nonincreasing expected returns.

The following is some notational shorthand: let $z_{-i} = (z_1, \cdots, z_{i-1}, z_{i+1}, \cdots, z_n)$; and let

$$E_{-i}(\cdot) = \int_0^1 \cdots \int_0^1 (\cdot) \prod_{j \neq i} g(z_j)\, dz_{-i}; \tag{3}$$

[8] The present model differs from that of Laffont and Tirole (1987) and McAfee and McMillan (1987b) in two respects. In those papers, several agents compete for the contract but only one is hired; here, there is no competition for contracts and the n agents work together in a team.

[9] In the case of no complementarity in production, $\mu = \Sigma_{i=1}^{n} y_i$, the team problem breaks up into n separate single-agent problems of the sort investigated by Laffont and Tirole (1986, 1987) and McAfee and McMillan (1987b).

(4) $$E_x(\cdot) = \int_0^\infty (\cdot) f(x, y)\, dx;$$

(5) $$V(y, z) = E_x U(x, y, z).$$

Define

(6) $$\gamma(y_i, z_i) = c(y_i, z_i) - \frac{1 - G(z_i)}{g(z_i)} c_z(y_i, z_i).$$

$\gamma(y_i, z_i)$ will turn out to be the cost as perceived by the principal: the sum of production cost (the first term on the right side (6)) and the cost of inducing the agent to reveal his private information about his ability (the second term, recalling that $c_z < 0$).

Assume individual rationality: any of the agents can refuse to participate if the contract offered by the principal gives him too little expected utility; that is,

(7) $$E_x E_{-i} \pi_i \geq 0.$$

We first characterize the best individual contribution to output by each agent that the principal can evoke subject to the asymmetric-information constraints. Then we exhibit a contract which succeeds in generating this individual contribution.

Define a rule relating individual contribution y_i to abilities z by

(8) $$y^*(z) = \arg\max_y V(y, z) - \sum_{j=1}^{n} \gamma(y_j, z_j).$$

We assume that the underlying functions U, c, and g are such that $y^*(z)$ is continuously differentiable. This ensures that problems of the sort discussed by Grossman and Hart (1983) do not arise.

The proofs of all of the following results are in the Appendix.

LEMMA 1. *Suppose the payment functions p_i satisfy*

(9) $$E_x E_{-i} \pi_i |_{z_i = 0} = 0$$

and evoke individual contributions $y^*(z)$, given z, where $y^*(z)$ is defined by (8). Then the p_i's maximize the principal's expected utility subject to individual rationality and incentive compatibility.

Thus the principal seeks effort levels that maximize his expected utility less total cost (production cost plus the principal's information costs). The best feasible level of effort, implicitly defined by (8), is too little (except for any agent with the highest possible ability level, $z_i = 1$). To see this, note that, in a full-information world, the principal would maximize $V(y, z) - \Sigma_{i=1}^n c(y_i, z_i)$. But, for $z_i < 1$, $\gamma_y(y_i, z_i) > c_y(y_i, z_i)$; thus (8) results in too little effort. The asymmetric information distorts the outcome. The principal perceives costs as being too high, as he faces not only production costs but also costs of inducing the agents to reveal their private information.[10]

The principal's aim, then, is to find a contract that induces correct revelation of the agents' private information, z, and induces effort levels consistent with $y^*(z)$. Note the multi-agent aspect of the contract-design problem: the incentive-compatibility conditions must take account of the fact that one agent's misrepresenting his ability will change the other agents' effort levels, which will in turn affect the first agent's payoff.

A payment function that satisfies the hypotheses of Lemma 1 need not exist. However, there are circumstances under which an especially simple payment function, linear in output, is optimal.

Define

$$\alpha_i(z) = \frac{c_y(y_i^*(z), z_i)}{\mu_i(y^*(z))}, \tag{10}$$

where $y^*(z)$ is defined by (8).

THEOREM 2. *If*

$$\frac{\partial \alpha_i(z)}{\partial z_i} \geq 0 \tag{11}$$

and

$$\frac{\partial y_i^*(z)}{\partial z_j} \geq 0 \qquad \text{for all } j \neq i,\ j = 1, \ldots, n, \tag{12}$$

then

$$p_i(x, z) = \alpha_i(z)[x - \mu(y^*(z))] + c(y_i^*(z), z_i) - \int_0^{z_i} c_z(y_i^*(s, z_{-i}), s)\, ds \tag{13}$$

is an optimal contract.

Thus, provided conditions (11) and (12) hold, the contract the principal offers each agent is linear in the team's output x. The sharing term is independent of the number of agents in the team, but does depend upon each team member's ability.

[10] To understand the role of c_z in the information cost (compare with equations (6) and (8)), notice that the cost function c is a reduced form; the corresponding structural form would explicitly involve the agent's effort. Consider one such structural form. Denote an agent's effort by e, and suppose his cost of effort is $K(e)$. Define a function $E(y, z)$ to be the minimal effort needed by an agent of type z to produce y. Then $c(y, z) = K(e(y, z))$, so $c_z = K'E_z$. Here K' is the marginal cost of effort, and E_z is the rate at which the effort necessary to produce y falls as the agent's ability z rises. Thus $-c_z$ measures the reduced cost to the agent of producing an output appropriate to a slightly lower-type agent. That is, $-c_z$ is the marginal benefit to the agent of imitating a lower type. Thus the larger $-c_z$ is, the more rents the agent earns from his private information.

With the sharing term as in (10), the agent equates his marginal cost of effort to his marginal return, which is the share he receives of his marginal output.[11]

What is the meaning of the sufficient conditions (11) and (12) for the linear contract to work? Both are incentive-compatibility requirements. Condition (12) is a complementarity statement: it implies that the higher one agent's ability is, the more effort the principal asks of the other agents. With (12), if agent j understates his ability, the other agents will be asked to work less hard, so that total output, and therefore agent j's payment, are smaller. From (13), $\alpha_i(z)$ is the fraction of marginal team output that agent i is allowed to keep. Condition (11) says that this share increases with the agent's reported ability. Thus, for example, a low-ability agent is not tempted to overstate his ability because he is then faced with a payment function that is relatively sensitive to his effort.

The expected payment to agent i is found by taking expectations over z_{-i} and x in (13). Since μ is the expected value of x, the first term on the right side of (13) becomes zero. The second term is the agent's cost. The remaining term, $E_{-i} - \int_0^{z_i} c_z(y_i^*(s, z_{-i}), s)\, ds$, is therefore agent i's expected profit, attributable to his private information about his ability. The following result follows from the proof of Theorem 2.

COROLLARY. *Each agent's rent is increasing both in his own ability and in the other agent's abilities.*

Hence, agents prefer to belong to teams whose members have high productivity.

Note that, from (8) and (10), conditions (11) and (12) are well-defined in terms of the model's primitives. It is not, however, clear from inspecting (8) and (10) what (11) and (12) mean in terms of the basic elements of the model. The next theorem provides a more restrictive, but more understandable, set of conditions under which linear contracts are optimal.

THEOREM 3. *Suppose $U(x, y, z) = x$, and*

$$c_{yzz} \geq 0; \tag{14}$$

$$\frac{c_{yz}}{c_y} \leq \frac{c_{yyz}}{c_{yy}} \leq \frac{g(z_i)}{1 - G(z_i)}; \tag{15}$$

$$\frac{\partial}{\partial z_i} \frac{c_{yz}(1 - G(z_i))}{c_y g(z_i)} \geq 0; \tag{16}$$

$$\frac{\partial}{\partial z_i} \frac{1 - G(z_i)}{g(z_i)} < 1. \tag{17}$$

Then the linear contract (13) is optimal.

The inequality (17) is the hazard-rate condition that commonly arises in adverse-

[11] The optimal payment schedule is written here as a function of actual abilities z, rather than reported abilities $\hat{z}$, since it has been designed to elicit correct revelation of abilities, so $z = \hat{z}$.

selection problems (McAfee and McMillan 1987a, p. 708). The inequality (16) requires monotonicity of the hazard rate weighted by the proportional rate of change of marginal cost of effort with respect to ability. Condition (15) states (since the first term is negative and the last is positive) that the curvature of the cost function neither increases nor decreases too rapidly with ability. Condition (14) requires that increases in ability reduce the marginal cost of effort at a diminishing rate. These conditions for a linear contract to work involve combinations of third derivatives of the cost function with hazard-rate conditions, and so are not in themselves very informative. A set of special cases that satisfy these conditions has the cost function $c(y, z)$ being linear or quadratic and (17) holding.

COROLLARY. *For the case in which the principal's return is proportional to total output,* $0 < \alpha_i(z) \leq 1$.

Hence no individual's marginal rate of payment is either negative or in excess of 100 percent.

For an example, suppose the principal cares only about outputs so that $U(x, y, z) = x$; abilities are distributed uniformly, so that $G(z_i) = z_i$ if $0 \leq z_i \leq 1$; the agent's cost function is $c(y_i, z_i) = y_i(1 - z_i)$; and expected team output is $\mu = \Pi_{j=1}^{n} y_j^{\beta}$, with $0 < \beta < 1/n$. With these functional forms, $\gamma(y_i, z_i) = 2y_i(1 - z_i)$. If we solve for $y^*(z)$ from (8), then (10) implies that $\alpha_i = 1/2$ for $i = 1, \dots, n$: remarkably, the share is independent of the agent's own ability or his fellow team-members' abilities. (As is clear from the general analysis, however, this is a knife-edge case.) Also, $\partial y_i^*/\partial z_j > 0$, so the hypotheses of Theorem 2 are satisfied and the linear contract is optimal. If the team has two members, each receives a half share of any output beyond a fixed target. With three or more team members, ex ante payments are made to the principal, who then pays out more than 100 percent of the marginal value of output.[12]

3. OPTIMAL CONTRACTS BASED ON INDIVIDUAL CONTRIBUTIONS

Suppose now we give the principal more information. The principal can costlessly monitor the individual contributions, so that he can pay agent i according to his contribution y_i as well as, like before, all of the reported abilities. We continue to assume, however, that the principal cannot observe the agent's ability z_i or cost $c(y_i, z_i)$: he cannot decompose an individual's contribution into effort and ability, so that ability remains the agent's private information. Note that now the principal's knowledge of total output x is redundant: x is effectively an imprecise measure of $(y_1, \dots, y_n)$, containing no extra relevant information.

THEOREM 4. *The principal's maximum expected utility when he monitors individual contributions is the same as when he observes only total output.*

[12] Gandal and Scotchmer (1989) apply these techniques to the analysis of cooperative research and development, and find conditions under which firms can be given incentives to do efficient amounts of investment.

Thus, perhaps surprisingly, there is no free-rider problem in this model. There is no gain to the principal from monitoring. The reason for this is most clearly seen in the case of no adverse selection, as explained in Section 1 above. Without adverse selection, an optimal no-monitoring contract pays each individual at the margin 100 percent of marginal team output; this achieves an efficient outcome, the same outcome as would be attained with monitoring and paying full individual marginal products. Adding adverse selection does not break this equivalence between contracts based on team output and contracts based on individual contributions, because the rents that an agent receives as a result of his private information are completely determined by the $y^*(z)$ function; and a given $y^*(z)$ can be induced by either monitored or nonmonitored contracts.

Can the optimum again be implemented using linear contracts?

THEOREM 5. *If*

$$\frac{\partial}{\partial z_i} c_y(y_i^*(z), z_i) \geq 0, \tag{18}$$

then

$$p_i^m(y_i, z) = c_y(y_i^*(z), z_i)(y_i - y_i^*(z)) + c(y_i^*(z), z_i) - \int_0^{z_i} c_z(y_i^*(s, z_{-i}), s)\, ds \tag{19}$$

is an optimal contract under monitoring.

The agent's equating his marginal cost of effort, c_y, to his share of his marginal output forces his contractual share, α_i, to be c_y, as in (19). Condition (18) says that this share must increase with ability. This is an incentive-compatibility condition: if payments became less sensitive to output as ability increased, low-ability agents would have an incentive to overstate their abilities. Condition (18), which ensures the optimality of linear contracts based on individual contributions, says that, with effort adjusted optimally, marginal production costs increase with ability. Two effects are operating. With output held constant, higher ability reduces marginal cost (by assumption). But an increase in ability increases the effort level that is optimal from the principal's point of view, and because of diminishing returns this increases marginal cost. Condition (18) assumes that the functional forms are such that the called-for increase in effort is large enough that the latter effect dominates.

As was discussed in Section 1, the advantage of contracts based on individual contributions over contracts based on total output is in the incentives facing the principal. The basis for this observation is the following theorem.

THEOREM 6. *If $U(x, y, z) = x$, then*

$$\sum_{i=1}^{n} c_y(y_i^*(z), z_i) y_i^*(z) \leq \mu(y^*(z)). \tag{20}$$

Decompose payments under the linear contracts into fixed and variable payments. Then (20) says that variable payments are smaller under the monitored contract than the expected output. The principal keeps some of the output for himself. The inequality (20) is strict if there are decreasing returns to scale in the team's production.

Sometimes (18) will fail to be satisfied and a contract linear in individual contribution, y_i, will not work even though one linear in team output, x, does. However, whenever a contract linear in x is optimal, a contract linear in a function of y_i will implement the optimum: in particular, $E_{-i}\mu(y_i, y^*_{-i}(z))$ works. (This is because the monitored contract is, by this function, in effect converted back to a nonmonitored contract.) Since there are no natural units for what we have been calling individual contribution to output, such a transformation changes nothing essential.

In the example given at the end of the last section, $c_y(y^*_i(z), z_i) = 1 - z_i$ is decreasing in z_i. Thus a monitored contract linear in individual contributions y_i is not feasible, even though, as we saw, a contract linear in team output is feasible and optimal. However, a monitored contract linear in y_i^β is optimal.

In the world, as opposed to in this model, inefficiencies are undoubtedly inherent in team production. Chinese agriculture, for example, became much more productive after the introduction of the responsibility system, under which individual farmers were remunerated according to their own outputs rather than, as before, according to the commune's output (McMillan, Whalley and Zhu 1989). What our model suggests is that the source of team problems is not the unobservability of team members' efforts or abilities per se. The source of team inefficiencies must be sought elsewhere, in features assumed away in our analysis: team members' risk aversion (Holmstrom 1982), or collusion among team members (Itoh 1989; Weiss 1988).[13]

4. CONCLUSION

This paper was written by a team. How should the principal, the chairman of our department, apportion merit pay to the agents, us, for our output, this article? Each author of an n-authored article usually gets $(1/n)$th of the credit. Our theory indicates that it is not in the chairman's interest to use this rule. If the chairman knew for sure our abilities, he would maximize our productivity by giving each of us 100 percent of the credit: he could compensate by setting base salaries lower than our alternative opportunities. If, however, the chairman were uncertain of our relative abilities to contribute to our joint work, he should give less than 100 percent credit to each of us; but the shares might still sum to more than 100 percent. In this case, before this article was written, the chairman should have offered each of us a menu of sharing arrangements, with higher shares associated with lower base salaries. By our selection of shares, we would each have revealed our own

[13] Mookherjee and Reichelstein (1990) show that the optimal individual contributions described by Lemma 1 can usually be implemented in dominant strategies; but the group-payment scheme of Theorem 5 typically cannot be implemented in dominant strategies. This, they suggest, provides a reason for monitoring.

estimates of our abilities. The more able coauthor would have chosen the higher share and thereby would have been induced to work harder.

University of Texas, Austin, U.S.A.
University of California, San Diego, U.S.A.

APPENDIX

PROOF OF LEMMA 1. Since i chooses y_i and $\hat{z}_i$ optimally, the Envelope Theorem yields

$$\text{(A1)} \qquad \frac{dE_{-i}\pi_i}{dz_i} = \left.\frac{\partial E_{-i}\pi_i}{\partial z_i}\right|_{\hat{z}_i = z_i} = -c_z(y_i, z_i) \geq 0.$$

Thus, individual rationality reduces to

$$\text{(A2)} \qquad E_x E_{-i}\pi_i|_{z_i=0} \geq 0.$$

Hence

$$\text{(A3)} \quad \Phi = E\left[U(x, y, z) - \sum_{i=1}^{n} p_i(x, z_i, z_{-i})\right]$$

$$= EU(x, y, z) - \sum_{i=1}^{n} E[\pi_i + c(y_i, z_i)]$$

$$= EE_x U(x, y, z) - E\sum_{i=1}^{n} c(y_i, z_i) - \sum_{i=1}^{n}\int_0^1 E_{-i}\pi_i g(z_i)\, dz_i$$

$$= E\left[E_x U(x, y, z) - \sum_{i=1}^{n} c(y_i, z_i)\right]$$

$$- \sum_{i=1}^{n}\left(E_{-i}(-\pi_i(1 - G(z_i)))\Big|_0^1 + \int_0^1 \frac{dE_{-i}\pi_i}{dz_i}(1 - G(z_i))\, dz_i\right)$$

$$= E\left[V(y, z) - \sum_{i=1}^{n} c(y_i, z_i)\right] + \sum_{i=1}^{n} E_{-i}\pi_i\Big|_{z_i=0}$$

$$- E_{-i}\int_0^1 c_z(y_i, z_i)(1 - G(z_i))\, dz_i$$

$$= E\left[V(y, z) - \sum_{i=1}^{n} c(y_i, z_i)\right]$$

$$+ \sum_{i=1}^{n} E_{-i} \int_0^1 c_z(y_i, z_i) \frac{1 - G(z_i)}{g(z_i)} g(z_i)\, dz_i$$

$$= E\left[V(y, z) - \sum_{i=1}^{n} \left[c(y_i, z_i) - c_z(y_i, z_i) \frac{1 - G(z_i)}{g(z_i)}\right]\right]$$

$$= EV(y, z) - \sum_{i=1}^{n} \gamma(y_i, z_i).$$

The third line uses Fubini's theorem, the fourth uses integration by parts, and the fifth optimizes by setting $E_{-i}\pi_i|_{z_i=0} = 0$. It follows immediately that $y^*(z)$ pointwise maximizes Φ. Q.E.D.

PROOF OF THEOREM 2. The profit of a type z_i agent reporting $\hat{z}_i$ and contributing y_i is

(A4) $$\pi_i(z_i, \hat{z}_i, y_i) = \alpha_i(\hat{z}_i, z_{-i})[\mu(y_i, y^*_{-i}(\hat{z}_i, z_{-i})) - \mu(y^*(\hat{z}_i, z_{-i}))]$$
$$+ c(y_i^*(\hat{z}_i, z_{-i}), \hat{z}_i) - c(y_i, z_i) - \int_0^{\hat{z}_i} c_z(y_i^*(s, z_{-i}), s)\, ds.$$

Now

(A5) $$\pi_{y_i} = \alpha_i(\hat{z}_i, z_{-i})\mu_i - c_y,$$

(A6) $$\pi_{y_i y_i} = \alpha_i \mu_{ii} - c_{yy} < 0.$$

Thus the agent reporting $\hat{z}_i$ chooses $\hat{y}_i$ satisfying

(A7) $$\alpha_i(\hat{z}_i, z_{-i}) = \frac{c_y(\hat{y}_i, z_i)}{\mu_i(\hat{y}_i, y^*_{-i}(\hat{z}_i, z_{-i}))}.$$

Thus

(A8) $$\frac{\partial \hat{y}_i}{\partial \hat{z}_i} = \frac{\dfrac{\partial \alpha_i}{\partial \hat{z}_i}(\hat{z}_i, z_{-i}) + \dfrac{c_y}{\mu_i^2} \sum_{j \neq i} \mu_{ij} \dfrac{\partial y^*_{-i}}{\partial \hat{z}_i}(\hat{z}_i, z_{-i})}{\dfrac{c_{yy}}{\mu_i} - \dfrac{c_y}{\mu_i^2}\mu_{ii}} \geq 0.$$

Note

(A9) $$\hat{y}_i|_{\hat{z}_i = z_i} = y_i^*(z).$$

To prove incentive compatibility, it is sufficient to prove

$$E_{-i}\pi_i(z_i, \hat{z}_i, \hat{y}_i) \le E_{-i}\pi_i(z_i, z_i, y_i^*(z)). \tag{A10}$$

However, this is implied by

$$E_{-i} \frac{\partial}{\partial \hat{z}_i} \pi_i(z_i, z_i, \hat{y}_i) = 0 \tag{A11}$$

and

$$\frac{\partial^2}{\partial \hat{z}_i \partial z_i} \pi_i(z_i, z_i, \hat{y}_i) \ge 0, \tag{A12}$$

since these imply

$$\frac{\partial}{\partial \hat{z}_i} E_{-i}\pi_i(z_i, \hat{z}_i, \hat{y}_i) \gtreqless 0 \text{ as } z_i \gtreqless \hat{z}_i. \tag{A13}$$

To prove (A11), it is sufficient to prove

$$\frac{\partial}{\partial z_i} \pi_i(z_i, \hat{z}_i, \hat{y}_i)\big|_{\hat{z}_i = z_i} = \frac{d}{dz_i} \pi_i(z_i, z_i, y_i^*(z_i)). \tag{A14}$$

$$\pi_i(z_i, z_i, y_i^*(z_i)) = -\int_0^{z_i} c_z(y_i^*(s, z_{-i}), s)\, ds. \tag{A15}$$

$$\frac{d\pi_i}{dz_i} = -c_z(y_i^*(z), z_i). \tag{A16}$$

$$\frac{\partial \pi_i}{\partial z_i}(z_i, \hat{z}_i, \hat{y}_i) = -c_z(\hat{y}_i, z_i). \tag{A17}$$

Thus

$$\frac{\partial \pi_i}{\partial z_i}(z_i, \hat{z}_i, \hat{y}_i)\big|_{\hat{z}_i = z_i} = \frac{d\pi_i}{dz_i}, \tag{A18}$$

as desired.

A proof of (A12) follows.

$$\frac{\partial^2 \pi_i}{\partial \hat{z}_i\, \partial z_i} = \frac{\partial}{\partial \hat{z}_i}[-c_z(\hat{y}_i, z_i)] = -c_{yz}\frac{\partial \hat{y}_i}{\partial \hat{z}_i} \ge 0. \tag{A19}$$

Q.E.D.

The proof of Theorem 3 will make use of the following lemma. (See, for example, Arrow and Hurwicz 1977, pp. 326, 371.)

LEMMA. *Suppose A is a negative semidefinite symmetric $n \times n$ matrix with $A_{ij} \geq 0$ if $i \neq j$. Then A^{-1} is composed entirely of nonpositive elements.*

PROOF OF THEOREM 3. To prove (12) note that

$$\gamma_y = c_y - \frac{1 - G(z_i)}{g(z_i)} c_{yz} \geq 0 \tag{A20}$$

and (using (15))

$$\gamma_{yy} = c_{yy} - \frac{1 - G(z_i)}{g(z_i)} c_{yyz} \geq 0. \tag{A21}$$

Thus γ is convex in y_i.

$$\gamma_{yz} = c_{yz}\left[1 - \frac{\partial}{\partial z_i} \frac{1 - G(z_i)}{g(z_i)}\right] - \frac{1 - G(z_i)}{g(z_i)} c_{yzz} \leq 0, \tag{A22}$$

since $c_{yzz} \geq 0$.

Let $\Gamma_y = (\gamma_y(y_1, z_1), \dots, \gamma_y(y_n, z_n))$, so that y^* satisfies, since $V = \mu$,

$$\mu_y(y^*(z)) - \Gamma_y(y^*(z), z) = \vec{0}. \tag{A23}$$

Thus

$$y^{*\prime}(z) = [\mu_{yy} - \Gamma_{yy}]^{-1}\Gamma_{yz}. \tag{A24}$$

Since μ is concave and γ is convex in y (by (A21)), $\mu_{yy} - \Gamma_{yy}$ is negative definite and has nonnegative off-diagonal elements. Thus, from the preceding lemma, $(\mu_{yy} - \Gamma_{yy})^{-1}$ has nonpositive elements. By (A22), Γ_{yz} is a diagonal matrix with nonpositive elements, so $y^{*\prime}(z)$ is composed of nonnegative elements, as desired.

The proof of (11) follows. Since $V(y, z) = \mu(y)$, y^* satisfies

$$\mu_i(y^*(z)) = \gamma_y(y_i^*(z), z_i) = c_y(y_i^*(z), z_i) - \frac{1 - G(z_i)}{g(z_i)} c_{yz}(y_i^*(z), z_i). \tag{A25}$$

Thus

$$\alpha_i(z) = \frac{c_y(y_i^*(z), z_i)}{\mu_i(y^*(z))} = \frac{1}{1 - \dfrac{1 - G(z_i)}{g(z_i)} \dfrac{c_{yz}(y_i^*(z), z_i)}{c_y(y_i^*(z), z_i)}} \leq 1. \tag{A26}$$

Thus

$$\text{(A27)} \qquad \operatorname{sgn}\left(\frac{\partial \alpha_i}{\partial z_i}\right) = \operatorname{sgn}\left(\frac{\partial}{\partial z_i}\,\frac{1-G(z_i)}{g(z_i)}\,\frac{c_{yz}(y_i^*(z), z_i)}{c_y(y_i^*(z), z_i)}\right)$$

$$= \operatorname{sgn}\left\{\frac{\partial}{\partial z_i}\,\frac{c_{yz}}{c_y}\,\frac{1-G(z_i)}{g(z_i)} + \frac{1-G(z_i)}{g(z_i)}\left(\frac{\partial}{\partial y_i}\,\frac{c_{yz}}{c_y}\right)\frac{\partial y_i^*(z)}{\partial z_i}\right\}.$$

So, using (15) and (16),

$$\frac{\partial \alpha_i}{\partial z_i} \geq 0. \qquad \text{Q.E.D.}$$

PROOF OF THEOREM 4. As in Theorem 1,

$$\text{(A28)} \qquad \frac{dE_{-i}\pi_i}{dz_i} = \frac{\partial E_{-i}\pi_i}{\partial z_i} = -c_z(y_i, z_i).$$

It follows that equation (A3) still holds (the proof is the same), and thus the principal obtains the same utility as before. Q.E.D.

PROOF OF THEOREM 5.

$$\text{(A29)} \qquad \pi_i(z_i, \hat{z}_i, y_i) = c_y(y_i^*(\hat{z}_i, z_{-i}), \hat{z}_i)[y_i - y_i^*(\hat{z}_i, z_{-i})] + c(y_i^*(\hat{z}_i, z_{-i})) - c(y_i, z_i) - \int_0^{\hat{z}_i} c_z(y_i^*(s, z_{-i}), s)\, ds.$$

$$\text{(A30)} \qquad \pi_{y_i} = c_y(y_i^*(\hat{z}_i, z_{-i}), \hat{z}_i) - c_y(y_i, z_i).$$

$$\text{(A31)} \qquad \pi_{y_i y_i} = -c_{yy} < 0.$$

Thus, an agent choosing $\hat{z}_i$ chooses $\hat{y}_i$ to satisfy

$$\text{(A32)} \qquad c_y(y_i^*(\hat{z}_i, z_{-i}), \hat{z}_i) = c_y(\hat{y}_i, z_i).$$

Thus,

$$\text{(A33)} \qquad \frac{d\hat{y}_i}{d\hat{z}_i} = \frac{\dfrac{\partial}{\partial \hat{z}_i} c_y(y_i^*(\hat{z}_i, z_{-i}), \hat{z}_i)}{c_{yy}(\hat{y}_i, z_i)} \geq 0.$$

It is easily seen that

$$\text{(A34)} \qquad \left.\frac{\partial \pi_i}{\partial z_i}\right|_{\hat{z}_i = z_i} = \frac{d\pi_i}{dz_i}(z_i, z_i, \hat{y}_i).$$

$$\text{(A35)} \qquad \frac{\partial}{\partial z_i}\pi(z_i, \hat{z}_i, \hat{y}_i) = -c_z(\hat{y}_i, z_i).$$

$$\text{(A36)} \qquad \frac{\partial^2}{\partial \hat{z}_i \, \partial z_i} \pi(z_i, \hat{z}_i, \hat{y}_i) = -c_{yz} \frac{\partial \hat{y}_i}{\partial \hat{z}_i} \geq 0,$$

forcing incentive compatibility. Q.E.D.

PROOF OF THEOREM 6.

$$\text{(A37)} \qquad \sum_{i=1}^{n} c_y(y_i^*(z), z_i) y_i^*(z) = \sum_{i=1}^{n} \alpha_i(z) \mu_i(y^*(z)) y_i^*(z)$$

$$\leq \sum_{i=1}^{n} \mu_i(y^*(z)) y_i^*(z) \leq \mu(y^*(z)).$$

The first inequality by (A26), the second by the concavity of μ. Q.E.D.

REFERENCES

ALCHIAN, A. A. AND H. DEMSETZ, "Production, Information Costs, and Economic Organization," *American Economic Review* 62 (1972), 777–795.

ARROW, K. J., "The Economics of Agency," in J. W. Pratt and R. J. Zeckhauser, eds., *Principals and Agents: The Structure of Business* (Boston: Harvard Business School Press), 37–51.

——— AND L. HURWICZ, *Studies in Resource Allocation Processes* (Cambridge: Cambridge University Press, 1977).

BAIMAN, S. AND J. S. DEMSKI, "Economically Optimal Performance Evaluation and Control Systems," *Journal of Accounting Research* 18 (1980), 184–220.

ESWAREN, M. AND A. KOTWAL, "The Moral Hazard of Budget-Breaking," *RAND Journal of Economics* 15 (1984), 578–581.

GANDAL, N. AND S. SCOTCHMER, "Coordinating Research through Joint Ventures," mimeo, University of California, Berkeley, October 1989.

GROSSMAN, S. J. AND O. D. HART, "An Analysis of the Principal-Agent Problem," *Econometrica* 51 (1983), 7–46.

GROVES, T., "Incentives in Teams," *Econometrica* 41 (1973), 617–631.

HOLMSTROM, B., "Moral Hazard in Teams," *Bell Journal of Economics* 13 (1982), 324–340.

ITOH, H., "Collusion, Incentives, and Risk Sharing," mimeo, University of California, San Diego, April 1989.

LAFFONT, J.-J. AND J. TIROLE, "Using Cost Observation to Regulate Firms," *Journal of Political Economy* 94 (1986), 614–641.

——— AND ———, "Auctioning Incentive Contracts," *Journal of Political Economy* 95 (1987), 921–938.

MCAFEE, R. P. AND J. MCMILLAN, "Auctions and Bidding," *Journal of Economic Literature* 25 (1987a), 699–738.

——— AND ———, "Competition for Agency Contracts," *RAND Journal of Economics* 18 (1987b), 296–307.

MCMILLAN, J., J. WHALLEY AND L. ZHU, "The Impact of China's Economic Reforms on Agricultural Productivity Growth," *Journal of Political Economy* 97 (1989), 781–807.

MOOKHERJEE, D. AND S. REICHELSTEIN, "Dominant Strategy Implementation of Bayesian Incentive Compatible Allocation Rules," mimeo, University of California, Berkeley, September 1990.

MYERSON, R. B., "Optimal Coordination Mechanisms in Generalized Principal-Agent Problems," *Journal of Mathematical Economics* 10 (1982), 67–81.

PICARD, P. AND P. REY, "Incentives in Cooperative Research and Development," mimeo, CEPREMAP, September 1987.

WEISS, A., "Incentives and Worker Behavior: Some Evidence," in H. R. Nalbantian, ed., *Incentives, Cooperation, and Risk Sharing* (New York: Rowman and Littlefield, 1988).