CHAPTER 16

Multitask principal–agent analyses: incentive contracts, asset ownership, and job design

BENGT HOLMSTROM AND PAUL MILGROM

Bengt Holmstrom is the Paul A. Samuelson Professor of Economics at the Massachusetts Institute of Technology. See also the chapter by Holmstrom and Roberts.

Paul Milgrom is the Shirley R. and Leonard W. Ely, Jr., Professor of Humanities and Science and Professor of Economics at Stanford University. See also the chapter by Milgrom and Roberts.

1. Introduction

In the standard economic treatment of the principal–agent problem, compensation systems serve the dual function of allocating risks and rewarding productive work. A tension between these two functions arises when the agent is risk averse, for providing the agent with effective work incentives often forces him to bear unwanted risk. Existing formal models that have analyzed this tension, however, have produced only limited results.[1] It remains a puzzle for this theory that employment contracts so often specify fixed wages and more generally that incentives within firms appear to be so muted, especially compared to those of the market. Also, the models have remained too intractable to effectively address broader organizational issues such as asset ownership, job design, and allocation of authority.

In this article, we will analyze a principal–agent model that (i) can account for paying fixed wages even when good, objective output measures are available and agents are highly responsive to incentive pay; (ii) can make recommendations and predictions about ownership patterns even when contracts can take full account of all observable variables and court enforcement is perfect;

Excerpted from Bengt Holmstrom and Paul Milgrom, "Multitask Principal–Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design," *Journal of Law, Economics, and Organization* 7 (1991): 24–52, with permission of Oxford University Press.

[1] Some of the predictive weaknesses of standard agency models are discussed in the surveys by MacDonald, Hart and Holmstrom, and Baker, Jensen, and Murphy.

https://doi.org/10.1017/CBO9780511817410.018

(iii) can explain why employment is sometimes superior to independent contracting even when there are no productive advantages to specific physical or human capital and no financial market imperfections to limit the agent's borrowings; (iv) can explain bureaucratic constraints; and (v) can shed light on how tasks get allocated to different jobs.

The distinguishing mark of our model is that the principal either has several different tasks for the agent or agents to perform, or the agent's single task has several dimensions to it. Some of the issues raised by this modeling are well illustrated by the current controversy over the use of incentive pay for teachers based on their students' test scores. Proponents of the system, guided by a conception very like the standard one-dimensional incentive model, argue that these incentives will lead teachers to work harder at teaching and to take greater interest in their students' success. Opponents counter that the principal effect of the proposed reform would be that teachers would sacrifice such activities as promoting curiosity and creative thinking and refining students' oral and written communication skills in order to teach the narrowly defined basic skills that are tested on standardized exams. *It would be better, these critics argue, to pay a fixed wage without any incentive scheme than to base teachers' compensation only on the limited dimensions of student achievement that can be effectively measured*.[2]

Multidimensional tasks are ubiquitous in the world of business. As simple examples, production workers may be responsible for producing a *high volume* of *good quality* output, or they may be required both to produce output and to care for the machines they use. In the first case, if volume of output is easy to measure but the quality is not, then a system of piece rates for output may lead agents to increase the volume of output at the expense of quality. Or, if quality can be assured by a system of monitoring or by a robust product design, then piece rates may lead agents to abuse shared equipment or to take inadequate care of it. In general, when there are multiple tasks, incentive pay serves not only to allocate risks and to motivate hard work, it also serves to direct the allocation of the agents' attention *among* their various duties. This represents the first fundamental difference between the multidimensional theory and the more common one-dimensional principal–agent models.

There is a second fundamental difference as well, and it, too, can be illustrated by reference to the problem of teaching basic skills: If the task of teaching basic skills could be separated from that of teaching higher-level thinking, then

[2] As a concrete illustration of the distortions that testing can cause, in 1989 a ninth-grade teacher in Greenville, South Carolina, was caught having passed answers to questions on the statewide tests of basic skills to students in her geography classes in order to improve her performance rating (*Wall Street Journal*, November 2, 1989).

these tasks could be carried out by different teachers at different times during the day. Similarly, in the example of the production worker, when the care and maintenance of a productive asset can be separated from the use of that asset in producing output, the problem that a piece-rate system would lead to inadequate care can be mitigated or even eliminated. In general, in multitask principal–agent problems, *job design is an important instrument for the control of incentives*. In the standard model, when each agent can engage in only one task, the grouping of tasks into jobs is not a relevant issue.

Our formal modeling of these issues utilizes our linear principal–agent model (Holmstrom and Milgrom, 1987), mainly specialized to the case where the agent's costs depend only on the *total* effort or attention the agent devotes to all of his tasks. This modeling assures that an increase in an agent's compensation in any one task will cause some reallocation of attention away from other tasks. First, we show that an optimal incentive contract can be to pay a fixed wage independent of measured performance, just as the opponents of incentives based on educational testing have argued. More generally, the desirability of providing incentives for any one activity decreases with the difficulty of measuring performance in any other activities that make competing demands on the agent's time and attention. This result may explain a substantial part of the puzzle of why incentive clauses are so much less common than one-dimensional theories would predict.

Second, we specialize our model to the case where the unmeasurable aspect of performance is how the value of a productive asset changes over time. The difficulties of valuing assets are well recognized, and the vast majority of accounting systems value assets using fixed depreciation schedules based on historical costs, deviating from this procedure only in exceptional circumstances. Under these conditions, when the principal owns the returns from the asset, the optimal incentive contract will provide only muted incentives for the agent to produce output, in order to mitigate any abuse of the asset or any substitution of effort away from asset maintenance. However, when the agent owns the asset returns, the optimal incentive contract will provide more intensive incentives to engage in production, in order to alleviate the reverse problem that the agent may use the asset too cautiously or devote too much attention to its care and improvement. This analysis supports Williamson's observation that "high-powered" incentives are more common in market arrangements than within firms, without relying on any assumptions about specific investments. Moreover, it provides a rudimentary theory of ownership, according to which the conditions that favor the agent owning the assets are (i) that the agent is not too risk averse, (ii) that the variance of asset returns is low, and (iii) that the variance of measurement error in other aspects of the agent's performance is low. Thus, it emphasizes measurement cost as an important determinant of

integration in contrast to the leading approaches, which stress asset specificity.[3]

. . .

Third, we explore how a firm might optimally set policies limiting personal business activities on company time. Again, it is not just the characteristics of the "outside activities" themselves that determine whether these activities should be permitted. We find that outside activities should be most severely restricted when performance in the tasks that benefit the firm – the "inside activities" – are hard to measure and reward. Thus, a salesperson whose pay is mostly in the form of commissions will optimally be permitted to engage in more personal activities during business hours than a bureaucrat who is paid a fixed wage, because the commissions direct the salesperson toward inside activities in a way that cannot be duplicated for the bureaucrat. Our theory also predicts that home office work should be accompanied by a stronger reliance on performance-based pay incentives, a prediction that seems to fit casual observation.

Our analysis of restrictions on outside activities underscores the fact that incentives for a task can be provided in two ways: either the task itself can be rewarded or the marginal opportunity cost for the task can be lowered by removing or reducing the incentives on competing tasks. Constraints are substitutes for performance incentives and are extensively used when it is hard to assess the performance of the agent. We believe this opens a new avenue for understanding large-scale organization. It also offers an alternative interpretation of the Anderson–Schmittlein evidence. It is inefficient to let a salesperson, whose performance is poorly measured, divert his time into commission selling of competing products. If the employer has an advantage in restricting the employees' other activities, as both Simon and Coase have argued, then problems with measuring sales performance will lead to employing an in-house sales force.

Finally, we obtain a series of results in the theory of job design, using a model in which the employer can divide responsibility for many small tasks between two agents and can determine how performance in each task will be compensated. The resulting optimization problem is a fundamentally nonconvex one, and we have had to make some extra assumptions to keep the analysis

[3] Alchian and Demsetz argued that monitoring difficulties account for the formation of firms, but their theory was subsequently rejected in favor of the view that asset specificity and *ex post* bargaining problems drive integration (Grossman and Hart; Williamson). We are reintroducing measurement cost as a key factor but in a way that differs from the original Alchian–Demsetz theory. In particular, we do not argue that owners can better monitor the workforce. Our approach is more closely related to Barzel's work.

tractable. Nevertheless, the results we obtain seem intriguing and suggestive. First, we find that each task should be made the responsibility of just one agent. To our knowledge, this the first formal derivation in the incentive literature of the principle of unity of responsibility, which underlies the theory of hierarchy. Second, we find that tasks should be grouped into jobs in such a way that the tasks in which performance is most easily measured are assigned to one worker and the remaining tasks are assigned to the other worker. This conclusion squares nicely with the intuition that it is the *differences* between the measurability of quantity and quality in production, or of the so-called basic skills and higher-order thinking skills in education, that make those incentive problems difficult. The theory indicates that even when the agents have identical *ex ante* characteristics, the principal should still design their jobs to have measurement characteristics that differ as widely as possible. The principal should then provide more intensive incentives and require more work effort from the jobholder whose performance is more easily measured.

Our results are variations on the general theme of second best, which stresses that when prices cannot allocate inputs efficiently, then optimal incentives will typically be provided by subsidizing or taxing all inputs. For instance, Greenwald and Stiglitz, in a vivid metaphor, point out the value of a government subsidy for home fire extinguishers, since homeowners with fire insurance have too little incentive to invest in all forms of fire prevention and to fight fires once they have started. This mechanism has been most extensively analyzed in the theory of optimal taxation and in welfare theory.

However, the study of interdependencies among incentives and the use of instruments other than compensation to alleviate incentive problems have entered agency analyses more recently. Lazear argues that where cooperation among workers is important, we should expect to see less wage differentiation, that is, "lower-powered" incentives. Holmstrom and Ricart i Costa have observed how a firm's capital budgeting policy, including the hurdle rate and the way the firm assesses idiosyncratic risks, can affect the willingness of risk-averse managers to propose risky investment projects. Milgrom and Milgrom and Roberts have studied how organizational decision processes affect the allocation of effort between politicking and directly productive work. Farrell and Shapiro show that a price clause may be worse than no contract at all, because it reduces incentives to supply quality; this is similar to our result that it may be optimal to provide no quantity incentives when quality is poorly measured.

Some articles containing related ideas have been developed contemporaneously. Itoh (1991), in an analysis complementary to ours, studies conditions under which an employer might induce workers to work separately on their tasks, and those in which it is best for them to spend some effort helping one another. Laffont and Tirole show that concerns for quality help explain the use of cost plus contracting in procurement. Baker investigates a model in which

observable proxies of marginal product are imperfect in a way that causes the agent to misallocate effort across contingencies and therefore leads to incentives that are not as powerful as standard theory would suggest. Minahan reports a result on task separation that suggests a job design similar to ours but based on a different argument, as we will later explain.

The remainder of this article is organized as follows. In Section 2, we recapitulate our basic principal–agent theory, upon which the entire analysis is based. In Section 3, we specialize the analysis to the case where the agent's costs depend only on the total attention supplied and prove the various propositions about the optimality of fixed wages, the factors determining the assignment of ownership, and the optimal limits on outside business activities. In Section 4, we consider restrictions on private tasks. In Section 5, we offer a summary and suggest directions in which this line of research can be taken.

. . .

3. Allocation incentives for effort and attention

3.1 The effort and attention allocation model

We now move to a group of models in which the agent's effort or attention is a homogeneous input that can be allocated among tasks however the agent likes. We shall suppose that effort in the various tasks is perfectly substitutable in the agent's cost function. More formally, we suppose that the agent chooses a vector $t = (t_1, \ldots, t_m)$ at a personal (strictly convex) cost $C(t_1 + \cdots + t_m)$, leading to expected profits $B(t)$ and generating signals $x(t) = \mu(t) + \epsilon$. Then, if the agent increases the amount of time or attention devoted to one activity, the marginal cost of attention to the other activities will grow larger.

Contrary to most earlier principal–agent models, we shall not suppose that all work is unpleasant. . . . A worker on the job may take pleasure in working up to some limit; incentives are only required to encourage work beyond that limit. Formally, we assume that there is some number $\bar{t} > 0$ such that $C(t) \leq 0$ for $t \leq \bar{t}$ and $C(\bar{t}) = 0$. This is important, because it means that contracts that provide for fixed wages may still elicit some effort, though more may be elicited by providing positive incentives. It also means that there is a range of effort allocations among which the agent is indifferent and willing to follow the principal's preference.

3.2 Missing incentive clauses in contracts

One of the most puzzling and troubling failures of incentive models has been their inability to account for the paucity of explicit incentive provisions in

actual contracts. For example, it is surprisingly uncommon in contracts for home remodeling to incorporate explicit incentives for timely completion of construction, even though construction delays arise frequently and can be profoundly disruptive to the homeowner. There can be little doubt that such clauses could be written into the contracts; similar clauses are common in commercial construction contracts. We shall argue that these facts can best be understood as a result of the greater standardization of commercial construction and the consequent ability of commercial buyers to specify and monitor quality standards. The innovation in our analysis is that our explanation of the presence or absence of the timely completion clause lies in an examination of the principal's ability to monitor *other aspects* of the agent's performance.[4]

Thus, suppose that some desirable attributes of the contractor's performance (such as courtesy, attention to detail, or helpful advice) are unmeasurable but are enhanced by attention t_1 spent on that activity, while other aspects of quality (such as timely completion) are measurable (perhaps imperfectly) and enhanced by attention t_2 devoted to this second activity. Supposing that the measured quality is one dimensional, we may write $\mu(t_1,t_2) = \mu(t_2), x = \mu + \epsilon$. As we have seen, the agent's efficient compensation contract pays an amount $S = \alpha x + \beta$.

Suppose that the overall value of the job to the homeowner is determined by the function $B(t_1,t_2)$. To model the idea that the first activity is "very important" and that both activities are valuable, we assume that B is increasing and that $B(0,t_2) = 0$, for all $t_2 \geq 0$.

Proposition 1. For the home contractor model specified in the last paragraph, the efficient linear compensation rule pays a fixed wage and contains no incentive component ($\alpha = 0$), even if the contractor is risk neutral.

. . .

The ideas that underlie this analysis have many applications. For example, piece rates are relatively rare in manufacturing and, where they are used, they are frequently accompanied by careful attention to monitoring the quality of the work. Our analysis indicates that if quality were poorly measured, it would be expensive or impossible to maintain good quality while using a piece-rate scheme. Similarly, where individuals spend part of their efforts on individual projects and part on team production, and assuming that individual contributions to the team effort are difficult to assess, it would be dangerous to provide incentives for good performance on the individual projects. The

[4] Another plausible explanation is that home construction contracts are frequently changed to reflect design modifications, and timely completion clauses would be nullified by these changes.

problem, of course, is that individuals may shift their attention from the team activity where their individual contributions are poorly measured to the better measured and well-compensated individual activity. For this reason, piece-rate schemes may be especially dysfunctional in large hierarchies.

3.3 "Low-powered incentives" in firms

A similar model can be used to explain Williamson's observation that the incentives offered to employees in firms are generally "low-powered" compared to the "high-powered" incentives offered to independent contractors. Like Williamson, we distinguish employees from independent contractors by the condition of asset ownership: Employees use and develop assets that are owned by others while contractors use and develop their own assets.

Once again, the heart of our modeling is our assumption that there are multiple activities to be undertaken and that the allocation of time and attention between them is crucial. Thus, let the expected gross profit from the enterprise be the sum of two parts, $B(t_1) + V(t_2)$, where B represents the expected net receipts and V the expected change in the net asset value. We assume that B and V are increasing, concave, and twice continuously differentiable and that $B(0) = V(0) = 0$. The actual change in asset value, $V + \epsilon_v$, accrues to whomever owns the asset. Assets are notoriously hard to value (that is why accountants generally use historical cost as a valuation basis), so we assume that there is no performance indicator for the asset enhancement activity t_2. The primary activity t_1 is to produce output for sale in the current period: its indicator is $x = \mu(t_1) + \epsilon_x$, where μ is increasing and concave. We assume that ϵ_x and ϵ_v are independent.

We consider two alternative organizational modes – *contracting*, in which the change in asset value accrues to the agent, and *employment*, in which the change in asset value accrues to the firm or principal. The crucial difference between these lies in the incentives for the agent to engage in the two kinds of activities. To focus on the most interesting case, we will assume that it is highly desirable to induce the agent to devote a positive amount of effort to both activities. Let

$$\pi^1 = \operatorname*{Max}_{t_1} B(t_1) - C(t_1),$$

$$\pi^2 = \operatorname*{Max}_{t_2} V(t_2) - C(t_2),$$

$$\pi^{12} = \operatorname*{Max}_{t_1} B(t_1) + V(\bar{t} - t_1) - C(\bar{t}).$$

Proposition 2. Assume that $\pi^{12} \geq \text{Max}(\pi^1, \pi^2)$. Then, the optimal employment contract always entails paying a fixed wage ($\alpha = 0$). Whenever the independent contracting relation is optimal, it involves "high-powered incentives" ($\alpha > 0$). Furthermore, there exist values of the parameters $r, \sigma_v^2, \sigma_x^2$ [ed.: risk aversion and the variances associated with V and x, respectively] for which employment contracts are optimal and others for which independent contracting is optimal. If employment contracting is optimal for some fixed parameters $(r, \sigma_v^2, \sigma_x^2)$, then it is also optimal for higher values of these parameters. Similarly, if independent contracting is optimal, then it is also optimal for lower values of these parameters.

. . .

[A] piece of evidence consistent with our model comes from the fast-food industry. Firms such as McDonald's and Burger King own about 30% of their stores and franchise the rest. The difference in incentives between franchisees and owner-managed firms is striking. Franchisees pay royalties that are at most 10% of sales, corresponding to at least a 90% commission, whereas managers of company-owned stores typically receive no explicit incentives either on profit or sales (Krueger, 1991; Brickley and Dark, 1987). The difference in incentives is all the more remarkable, considering how similar the two types of stores are in all other aspects. According to our theory, the discontinuous shift in residual returns [$V(t_2)$] associated with franchising and the attendant shift in attention toward long-term asset values and cost containment, forces the franchise contract to increase short-term incentives sharply. Or, looked upon the other way, short-term incentives for employed managers must be muted to prevent them from allocating their attention away from important, but hard to measure, asset values.

4. Limits on outside activities

Our previous analysis emphasizes the importance of studying the full range of the agent's activities for analyzing incentives. If activities interact in the agent's cost function, incentive strength can be predicted only once the agent's whole portfolio of tasks is known. An equally important implication is that the principal can influence the agent's incentives by choosing the agent's portfolio of tasks. In the next section, we will study the optimal allocation of tasks between two agents. In this section, we consider how the principal might try to manage the agent's access to outside (private) activities.

Even casual observation makes it clear that the rules governing outside activities depend on the job. It is a commonplace observation that employees

in "responsible positions" are allowed more freedom of action than other employees and that they use that freedom in part to pursue personally beneficial activities. To analyze the issues that this observation raises, we begin with the assumption that it is easier for an employer to exclude an activity entirely than to monitor it and limit its extent. For example, a rule against personal telephone calls during business hours is found in many offices and seems to be motivated in part by its ease of enforcement compared, say, to a rule that limits the percentage of business hours devoted to personal calls to 2%. Although generalizations about employment all seem to have exceptions, a common feature of employment contracts is that the employer has authority to restrict the employee's outside activities during business hours, and sometimes after hours as well.

Assume then that the agent has a finite pool $K = \{1, \ldots, N\}$ of potential activities, which the principal can control only by exclusion. The returns to these tasks, which we will refer to as the agent's *personal business* for short, are assumed nonstochastic and to benefit the agent alone (in principle, these tasks could benefit the principal, too, but the analytics would be more complicated). The principal controls the agent's personal business by allowing the agent to engage only in a subset of tasks $A \subset K$. Within the *set of allowable tasks, A*, the agent can engage in as much or as little personal business as he pleases, but none outside A. To focus on the interactions between the agent's workplace activities and personal business, we represent workplace activities simply as a single task in which performance is imperfectly measured.

. . .

[Proposition 4.] (ii) If it becomes easier to measure the agent's performance (σ^2 decreases), or the agent becomes less risk averse (r decreases), then the agent's marginal reward will be raised and his personal business activities will be less curtailed.

. . .

[Thus], there will be more constraints on an agent's activities in situations where performance rewards are weak because of measurement problems. The rigid rules and limits that characterize bureaucracy, in this view, constitute an optimal response to difficulties in measuring and rewarding performance. Among the "personal business" activities that bureaucracies try to limit are collusion (Tirole; Holmstrom and Milgrom, 1990; Itoh, 1989) and influence activities (Milgrom; Milgrom and Roberts). The restrictions on trade between employees that Holmstrom and Milgrom (1990) recommend and the restrictions on communications that Milgrom and Roberts propose are examples of optimal

exclusion of activities that would be permitted or perhaps even encouraged in a first-best world.

. . .

5. Allocating tasks between two agents

In the single-agent model, the commission rates α_i serve three purposes: they allocate risk, motivate work, and direct the agent's efforts among his various activities. A trade-off arises when these objectives are in conflict with each other: Optimal risk sharing may be inconsistent with motivating work, and motivating hard work may distort the agent's allocation of efforts across tasks. Among the instruments available to the principal to alleviate these problems are job restructuring and relative performance evaluation: The former allows the principal to reduce the distortions in how attention is allocated among activities, while the latter enables the principal to lower the cost of incentives by using a more sensitive measure of actual performance.

5.1 Optimal groupings of tasks into jobs

Here we initiate the study of how incentive considerations might affect the grouping of tasks into jobs. We use a model that eliminates other important effects, such as differences among the agents and complementarities among task assignments. There are two identical agents, indexed $i = 1,2$, who allocate their attention across a continuum of tasks indexed by $k \in [0,1]$. Let $t_i(k)$ denote the attention agent i devotes to task k. We assume that the two agents can share a task and that their labor inputs are perfect substitutes. Thus, profit $B(t)$ is a function of the total time vector $t \equiv \{t(k): k \in [0,1]\}$, where $t(k) = t_1(k) + t_2(k)$. Likewise, the performance signal from task k, $\mu(t(k),k)$, only depends on the total attention $t(k)$ devoted to it. The error variance of task k is $\sigma^2(k) > 0$ and the errors are assumed independent.

. . .

Proposition 5. In the model described above, it is never optimal for the two agents to be jointly responsible for any task k.

. . .

This proposition reflects our earlier observation that providing incentives for an agent in any task incurs a fixed cost as the agent assumes some nontrivial fraction of the risk associated with that task (or its measurement). Since

we have assumed that the tasks are small relative to the agent's capabilities, assigning joint responsibility for any task would incur two fixed costs unnecessarily. As the proof demonstrates, if one begins with an arrangement in which some tasks are shared, it is possible to split the same tasks among the agents without affecting either the total effort required of either agent or the total effort allocated to any task. This rearrangement makes it possible to eliminate some of each agent's responsibilities [setting $\alpha_i(k) = 0$], thereby reducing the risk that the agent must bear and so increasing the total surplus of the three parties.

Having established that each task will be assigned to just one employee, we next turn to the issue of how the tasks will be grouped. With this in mind, it is convenient to redefine our variables. We reinterpret $\alpha_i(k)$ to be the *hypothetical* commission rate that the principal would need to pay in order to elicit the desired level of effort $t(k)$ from agent i if he were assigned task k.

. . .

Proposition 6. Suppose that the two agents devote different amounts of total attention to their tasks (i.e., $\bar{t}_1 \leq \bar{t}_2$). Then, tasks are optimally assigned in this model so that all the hardest-to-monitor tasks are undertaken by agent 1 and all the easiest-to-monitor tasks are undertaken by agent 2. That is, agent 1 is assigned all the tasks k for which $\rho(k) \geq \rho$, and agent 2 is assigned all those with $\rho(k) < \rho$, where ρ is defined in (23) [ed.: as an indicator of task measurability].

. . .

These results provide, in purely incentive-theoretic terms, an account of how activities might be grouped, with some employees specializing in activities that are hard to monitor and others in activities that are easily monitored. Separating tasks according to their measurability characteristics [$\rho(k)$] allows the principal to give strong incentives for tasks that are easy to measure without fearing that the agent will substitute efforts away from other, harder-to-measure tasks. The present model oversimplifies these issues by assuming that there are no restrictions on how the principal may group tasks. In the case of piece rates discussed in Section 3, it might not be possible to separate the tasks of providing high output from those of providing high quality: The worker might always be able to substitute speed for attention to details. Nevertheless, the results of Proposition 6 are suggestive.

. . .

6. Conclusion

The problem of providing incentives to agents and employees is far more intricate than is represented in standard principal–agent models. The performance measures upon which rewards are based may aggregate highly disparate aspects of performance into a single number and omit other aspects of performance that are essential if the firm is to achieve its goals. Commonly, the principal–agent problem boils down to this: Given a highly incomplete set of performance measures and a highly complex set of potential responses from the agent, how can the agent be motivated to act in the social interest?

Our approach emphasizes that incentive problems must be analyzed in totality; one cannot make correct inferences about the proper incentives for an activity by studying the attributes of that activity alone. Moreover, the range of instruments that can be used to control an agent's performance in one activity is much wider than just deciding how to pay for performance. One can also shift ownership of related assets, vary restrictions on the ways a job can be done, vary limits and incentives for competing activities, group related tasks into a single job, and so on.

In a related article (Holmstrom and Milgrom, 1991b), we study the simultaneous use of various instruments for controlling agents to derive new, testable results from the theory of organization. Our emphasis there is on how cross-sectional variations in the parameters that determine the optimal design of jobs, the optimal intensity of incentives, and the optimal allocation of ownership lead to covariations among endogenous variables that are similar to the patterns we find in actual firms.

Most past models of organization focus only on one instrument at a time for determining incentives and a single activity to be motivated. Newer theories, such as ours, that explicitly recognize connections between instruments and activities, offer new promise to explain the richer patterns of actual practice.