Journal of Economic Perspectives—Volume 12, Number 4—Fall 1998—Pages 115–132

Incentives in Organizations

Robert Gibbons

In 1975, Steven Kerr published "On the Folly of Rewarding A, While Hoping for B." The argument was simple: you get what you pay for. Kerr distilled this unifying theme from a disparate set of examples involving politicians, soldiers, doctors, orphanage directors, professors, and students, as well as manufacturing and clerical employees and even human-resource managers. From these examples, Kerr (pp. 779–80) concluded that two main causes of distorted incentives are "fascination with an 'objective' criterion, [where] individuals seek to establish simple, quantifiable standards against which to measure and reward performance" and "overemphasis on highly visible behaviors, [when] some parts of the task are highly visible while others are not." It took agency theory 15 years to express Kerr's title, not to mention to evaluate or extend his conclusions. During this period, agency theory was obsessed with the trade-off between incentives and insurance, even though clear-eyed observations like Kerr's about the design and performance of real incentive contracts suggested that several other issues are at least as important. Fortunately, recent work has brought agency theory not only to Kerr's position but beyond.

In this paper I summarize four new strands in agency theory that help me think about incentives in real organizations. As a point of departure, I begin with a quick sketch of the classic agency model. I then discuss static models of objective performance measurement that sharpen Kerr's argument; repeated-game models of subjective performance assessments; incentives for skill development rather than simply for effort; and incentive contracts between versus within organizations. I con-

■ *Robert Gibbons is Sloan Distinguished Professor of Management, Sloan School, Massachusetts Institute of Technology, and Research Associate, National Bureau of Economic Research, both in Cambridge, Massachusetts. His e-mail address is ⟨rgibbons@mit.edu⟩.*

clude by suggesting two avenues for further progress in agency theory: better integration with organizational economics, as launched by Coase (1937) and reinvigorated by Williamson (1975, 1985), and cross-pollination with other fields that study organizations, including industrial relations, organizational sociology, and social psychology.

The Classic Agency Model: Incentives versus Insurance

To set the stage, I begin with a brief description of the classic model in agency theory. By treating this model briefly, I do not mean to imply that the tradeoff between incentives and insurance is irrelevant, only that it is not nearly as central as it was once deemed.[1] In describing this and subsequent models, I omit some assumptions, derivations, and results; Prendergast (forthcoming) offers thorough introductions to many of these models.

Consider an agent who takes an unobservable action a to produce output y. For example, the production function might be linear, $y = a + \varepsilon$, where ε is a noise term. The principal owns the output but contracts to share it with the agent by paying a wage w contingent on output. For example, the wage contract might be linear, $w = s + by$, where the intercept s is the salary and the slope b is the bonus rate. The agent's payoff is $w - c(a)$, the realized wage minus the disutility of action $c(a)$. The principal's payoff is $y - w$, the realized output net of wages.

The key idea in this model is that the agent is risk-averse. (The principal may be as well.) A higher bonus rate b thus creates stronger incentives for the agent, but also imposes more risk on the agent. The extreme case of $b = 0$ offers the agent full insurance but creates no incentives; the other extreme, $b = 1$, gives the agent full title to the output y but offers the agent no insurance at all. The efficient bonus rate is between zero and one, depending on factors such as the amount of risk in ε and the parties' risk-aversions.

My perspective on risk in incentive contracting is nicely illustrated by work on sharecropping. The economic historians Lee Alston and Robert Higgs analyze three standard sharecropping contracts: wage labor, which imposes no risk on the agent ($b = 0$); crop sharing, which shares risk between the principal and the agent ($0 < b < 1$); and fixed-payment land rental, which leaves the agent with all the crop risk ($b = 1$). Higgs (1973) uses cross-sectional data on the southern United States for 1910 and finds that counties with greater crop risk made more use of risk-sharing; that is, more use of fixed wages and crop sharing rather than land rental. But Alston and Higgs (1982) show that there is enormous variation within each of these three main classes of contracts. For example, sharing contracts might apply to individuals, families,

[1] To be fair, there were several early models that studied incentive problems under risk-neutrality, such as delayed payments (Becker and Stigler, 1974; Lazear, 1979), tournaments (Lazear and Rosen, 1981), career concerns (Holmström, 1982) and efficiency wages (Shapiro and Stiglitz, 1984). But none of these models animated the field the way the classic model did.

or squads, and might be coupled with restrictions on other activities (such as on one's own plot) or asset-use opportunities (such as borrowing the landlord's mule for private purposes). Furthermore, Alston and Higgs also find significant variation in the use of the three classes of contracts even after controlling for risk. For example, different hands on the same plantation might be paid differently; "on some plantations one may find a dozen squads, each working on a different plan" (p. 330). Thus, the tradeoff between incentives and insurance has some explanatory power, but a great deal is hiding in the unexplained variation.

Similar evidence has been developed regarding cash compensation for chief executive officers of large U.S. firms. Aggarwal and Samwick (forthcoming) show that the classic tradeoff between insurance and incentives has some explanatory power. But Kole (1997) and Murphy (1998) document enormous variation in contract forms, suggesting that the tradeoff between incentives and insurance is again far from all that matters.

Objective Performance Measurement

I find the classic agency model strangely distant from real attempts to tie pay to performance, such as those described by Kerr (1975). The following examples, drawn from a seemingly infinite supply, suggest that Kerr's observations on the folly of rewarding A while hoping for B remain true today (Baker, Gibbons and Murphy, 1994, p. 1125):

> At the H.J. Heinz Company, division managers received bonuses only if earnings increased from the prior year. The managers delivered consistent earnings growth by manipulating the timing of shipments to customers and by prepaying for services not yet received, both at some cost to the firm. At Dun & Bradstreet, salespeople earned no commission unless the customer bought a larger subscription to the firm's credit-report services than in the previous year. In 1989, the company faced millions of dollars in lawsuits following charges that its salespeople deceived customers into buying larger subscriptions by fraudulently overstating their historical usage. In 1992, Sears abolished the commission plan in its auto-repair shops, which paid mechanics based on the profits from repairs authorized by customers. Mechanics misled customers into authorizing unnecessary repairs, leading California officials to prepare to close Sears' auto-repair business statewide.

There are of course exceptions, which I interpret as proving Kerr's (1975) rule. For instance, Lazear (1997) describes the transition from salaries to piece rates at a firm that installs auto windshields. The data show convincingly that output increased, due to two predicted effects: piece rates provided stronger incentives for hard work and also induced the self-selection of a workforce that benefited from those incentives. Unfortunately, not all jobs are as narrowly defined and easily mon-

itored as windshield installation; more typically, organizations get what they pay for. For two recent examples based on large micro datasets rather than the case studies above, see Brown, Harlow and Starks (1996) and Chevalier and Ellison (1997) on how the convex relationship between fund performance and assets under management caused risk-taking portfolio choices by ostensibly conservative mutual funds, and Anderson, Burkhauser and Raymond (1993) and Cragg (1997) on how the Job Training Partnership Act rewarded training providers for re-employment outcomes rather than for value added, so providers chose participants who were likely to find jobs even without training.

In this section I describe three static models in which firms get what they pay for: Baker (1992), Lazear (1989), and Holmström and Milgrom (1991). The key innovation in these models is to reject a strong but unremarked assumption in the classic agency model, where y is called "output," as though it could easily be measured. This label is misleadingly simple: in the classic model y reflects everything the principal cares about, except for wages (that is, the principal's payoff is $y - w$). Therefore, I henceforth call y the agent's "total contribution to firm value," to emphasize that it encompasses all the agent's actions (including mentoring, team production, and so on) and all the effects of these actions (both long- and short-run). In many settings, it is very difficult to measure synergies or sabotage across agents and/or very difficult to predict the long-run consequences of an agent's actions based on the observed short-run contribution. To analyze such settings, I henceforth assume that no contract based on y can be enforced in court, including but not limited to the linear contract $w = s + by$.

Of course, even when contracts based on y are not available, other contracts can be enforced in court. Such contracts are based on alternative performance measures—such as the number of units produced, with limited adjustment made for quality, timely delivery, and so on. Let p denote such an alternative performance measure; the wage contract might then be linear, $w = s + bp$. As in the classic agency model, a large value of b will create strong incentives, but now the agent's incentives are to produce a high value of p, not of y.

To illustrate one of Baker's (1992) main themes, suppose that there are two kinds of actions the agent can take, a_1 and a_2. In this setting, the contract $w = s + bp$ creates incentives that depend on the bonus rate b and on the way the actions a_1 and a_2 affect the performance measure p. But the marginal social benefits of the agent's actions depend on how a_1 and a_2 affect the agent's total contribution to firm value, y. To induce the agent to choose first-best actions, a contract must create incentives that match the marginal social benefits. But Baker argues that this is often impossible. As a trivial example, suppose that p is the sum of a_1 and a_2 but that y is the sum of a_1 and twice a_2. In a broad class of such examples, no contract can cause the agent's incentives to match the marginal social benefits of the agent's actions. Baker also explores the role of uncertainty, such as where the effects of a_1 on p and on y are uncertain and imperfectly correlated. Again, in a broad class of such examples, no contract can cause the agent's incentives to match the marginal social benefits of the agent's actions.

Lazear (1989) emphasizes a related theme: weak incentives may be more efficient than strong but dysfunctional incentives. He extends Lazear and Rosen's (1981) tournament model by allowing the agents to use two kinds of actions in attempting to win the tournament: effort, as in the original model, and now also sabotage. A big prize for winning the tournament induces not only a great deal of effort but also a great deal of sabotage, so the efficient prize level is smaller when sabotage is possible than when effort is the only action agents can use in attempting to win the tournament. Baker (1992) reaches a similar conclusion: it is no use creating strong incentives for the wrong actions. Indeed, the Baker and Lazear models are fundamentally similar—the performance measure p for the single agent in Baker's model is analogous to the relative performance of one agent over the other in Lazear's tournament; in both models, incentives are distorted because it is impossible to contract on a single agent's total contribution to firm value.

Holmström and Milgrom (1991) develop several other models in this "multitask" spirit. In some of their models, measured performance omits important dimensions of total contribution; for example, it may be that action a_1 contributes both to the performance measure p and to the total contribution y, but action a_2 contributes only to y, and does not affect p at all. Naturally, a contract that ties wages to such a performance measure will induce the agent to ignore the action a_2, even if a_2 greatly enhances the agent's total contribution to firm value. Holmström and Milgrom explore other models in which the reverse happens: action a_1 contributes to both p and y, but action a_2 contributes only to the performance measure and does not affect total contribution at all. In this case, of course, tying wages to such a performance measure creates an incentive for the agent to take action a_2, even if a_2 is irrelevant to or greatly reduces the agent's total contribution. Lazear's model has this latter form, as does the following Holmström-Milgrom model concerning restrictions on the activities permitted on the job.

To see why job restrictions may be an important part of an incentive plan, suppose that measured performance is based only on action a_1, but that the agent receives private value based on action a_2. For example, one could build a model in which a_2 increases the agent's visibility to other employers and hence increases the agent's market value, but does not change the agent's measured performance or total contribution at the present employer. Suppose also that the two actions compete for the agent's attention, in the sense that increasing the level of one action increases the marginal cost of the other. The problem in this setting is that the contractual incentives for the first action have to compete with the private attractions of the second. For plausible parameterizations, low values of the bonus rate b cause the agent to ignore action a_1 entirely, focusing solely on the private gains from taking action a_2. But if the principal could define the job in such a way as to exclude a_2 from the activities allowed on the job, then the agent's incentive to take the first action would depend on the bonus rate b in the usual way.

Using job restrictions in combination with an incentive plan illustrates the more general idea of using multiple instruments to provide a balanced package of incentives. Cockburn, Henderson and Stern (1998) present interesting evidence in this spirit,

concerning the management of researchers in the pharmaceutical industry. Pharmaceutical firms need such researchers to do two things: generate immediately useful output (such as drug patents) and invest in fundamental knowledge for the longer term (by attending conferences, publishing papers, and pursuing basic research). Research managers may balance these competing incentives by using the internal capital market to reward the former (by providing more funding to groups that produce more patents) and promotion policies to reward the latter (by promoting researchers who are active in the scientific community). Consistent with this hypothesis, Cockburn, Henderson and Stern find that research programs that offer strong incentives on one of these dimensions also offer strong incentives on the other.

The lessons to this point are that: 1) objective performance measures typically cannot be used to create ideal incentives; 2) efficient bonus rates are consequently often small; and 3) in multi-task settings, it is often helpful to use multiple instruments to provide a balanced package of incentives, and useful instruments range from direct cash payments to indirect organizational policies such as promotion criteria and job design. But the idea underlying all this—that measured performance differs from total contribution—also motivates a more direct approach to be taken up in the next section: paying agents for their total contributions by using subjective performance assessments rather than (only) objective performance measures.

Subjective Performance Assessments

Subjective assessments play important roles in many incentive contracts. Even foreign-exchange traders, whose books are marked to market at the end of each trading day, can have their incentive compensation tied to subjective judgments by their managers and co-workers. At Citicorp, for example, part of a trader's job is to execute orders that the bank's salespeople receive from clients, so part of the trader's bonus is based on subjective assessments by the salespeople concerning the timing and terms on which the trader executed orders.[2] Similarly, Lincoln Electric is well-known for its use of piece-rate formulas that tie a worker's pay to that worker's output, but about half of a worker's compensation is a bonus based on the supervisor's subjective assessment of the worker's cooperation, innovation, dependability, and so on (Fast and Berg, 1975). Moving beyond case studies, Hayes and Schaefer (1997) report evidence consistent with the use of subjective assessments when boards of directors decide the salary and bonus of chief executives: variation in the executive's current cash compensation that is not explained by current performance measures (such as stock return, sales, and earnings) predicts future variation in these performance measures. Finally, moving from microdata to everyday experience, subjective assessments of current performance may play a crucial role in determining future compensation, promotions, and contin-

[2] Personal communication in December 1997 from Julian Simmonds, Head of Foreign Exchange, Citicorp. See also Eccles and Crane (1988) for a similar description of the compensation of investment bankers.

ued employment, even where current compensation does not involve incentive pay of any kind. I cannot muster direct evidence of the role of subjective performance assessments in these future compensation, promotion, and employment decisions, but it seems uncontroversial that these decisions are not specified in advance as a function of objective performance measures.

In this section I discuss repeated-game models of "relational" incentive contracts—that is, agreements enforced by the parties' concerns for their reputations, as opposed to formal contracts enforced by a court. The advantage of relational contracts is that they can be based on (certain kinds of) subjective assessments, whereas formal contracts can be based only on objective measures. The disadvantage is that the parties can renege, so a relational contract must avoid creating (net) incentives for the parties to renege. I sketch a simple model based on Bull's (1987) early work; Malcomson (1998) surveys the subsequent literature.

Suppose that in each period of an ongoing employment relationship—which can be thought of formally as a repeated game—a worker chooses an unobservable action that influences that worker's total contribution to firm value for that period. As motivated in the previous section, I assume that the worker's contribution is too complex and subtle to be verified by an outsider, and so cannot be the basis of a contract enforced by a court. But even if the worker's contribution cannot be measured objectively, it often can be assessed subjectively by superiors who are well placed to observe the subtleties of the worker's behavior and opportunities. Following most of the literature, I will make the strong assumption that both the firm and the worker observe the worker's contribution perfectly. One phrase sometimes used to describe this assumption is that the worker's contribution is "observable but not verifiable;" that is, the worker's contribution is observable by the parties but not verifiable by outsiders such as a court. Relying heavily on this assumption, I will sketch a model in which the firm promises to pay the worker a bonus based on the worker's contribution. The analysis would be essentially unchanged if the firm and the worker both observed an imperfect proxy for the worker's contribution and the firm promised to pay a bonus based on this proxy (Baker, Gibbons and Murphy, 1994). But the analysis would be very different (and much more problematic) if the parties attempted to implement a relational contract based on a subjective assessment made by one party but observed only imperfectly by the other (Compte, 1998; Kandori and Matsushima, 1998).

To be concrete, suppose that the worker's total contribution to firm value is either High ($y = H$) or Low ($y = L$), and that higher levels of the worker's action increase the probability that the High contribution occurs. Imagine that the firm offers the worker a compensation package consisting of a base salary s paid when the worker accepts the offer and a relational-contract bonus B meant to be paid if the High contribution is achieved. In a single-period employment relationship in this setting, the firm would have no incentive to pay such a bonus. But in an ongoing relationship (with this or other workers), the firm's concern for its reputation may induce it to honor this relational contract.

In analyzing this repeated game, I focus on trigger strategies; roughly speaking, the parties begin by cooperating and then continue to cooperate unless one side

defects, in which case they refuse to cooperate forever after. In the context of this employment relationship, cooperating for the firm means paying the appropriate bonus and yields the firm some expected profit $E\pi(s, B)$ per period. Suppose that if the firm reneges on a bonus in a given period then no worker will trust the firm to pay a bonus in the future, so the firm will then receive a low payoff (normalized to zero) in every future period. The firm therefore can either pay the bonus today and receive $E\pi(s, B)$ forever after or not pay the bonus today and earn zero forever after. The firm will choose to pay the bonus if the present value of increased future profits from paying it exceeds the cost of paying the bonus today.[3]

This formulation yields intuitive comparative statics. For example, for low interest rates, the firm is patient and the prospect of future profits makes it willing to pay even a large bonus this period. Thus, for low interest rates, first-best incentives can be created through the choice of an appropriately high value of the bonus B. But for moderate interest rates, the present value of the expected profit from creating first-best incentives is not sufficiently high to make the firm willing to pay the high bonus necessary to create first-best incentives. Thus, for moderate interest rates, the firm can offer only a moderate bonus; the worker's incentive falls, and so expected profits fall, but the present value of future profits remains large enough to induce the firm to pay this moderate bonus when the worker produces a High contribution. Finally, for high interest rates, the present value of future profits is so small that the firm is not willing to pay even a moderate bonus. In fact, there may be no value of the bonus low enough that it is worthwhile for the firm to pay it, because a small bonus creates only a small incentive for the worker, and hence only a small present value of future profits for the firm. Thus, for high interest rates, the bonus must be either small or nonexistent.

Many well-known relational contracts have come under substantial stress (and sometimes failed) when the world has changed important parameters, such as the expected profit for the firm. For example, for several decades IBM made a "no layoffs" pledge to its employees. This was not a formal contract, enforceable by a court, but it was part of "the deal" at IBM: a shared understanding between the firm and its employees about how employment would proceed. As innovations in personal computers and workstations reduced the demand for mainframe computers, however, one could imagine that the value to IBM of living up to this pledge fell. Eventually, IBM abandoned the policy. A similar story can be told about bonus payments to investment bankers at First Boston in the early 1990s (Stewart, 1993). Many bankers left the firm after a second consecutive year in which they claimed that bonuses were unexpectedly low. The new parent company, Credit Suisse, claimed that bonuses were low because performance was low (and First Boston had

[3] Formally, paying the bonus yields a current profit of $H - s - B$ and an expected profit of $E\pi(s, B)$ in every future period, whereas not paying the bonus yields the larger current profit of $H - s$ but zero profit in every future period. For an interest rate of r, the present value of \$1 to be received in every future period is $\$1/r$, so the firm prefers to pay the bonus if $(H - s - B) + (1/r) \cdot E\pi(s, B) > (H - s) + (1/r) \cdot 0$, or $E\pi(s, B) > rB$.

indeed performed worse than its competitors), but many bankers argued that they should be paid the same as bankers at other firms. There may have been a legitimate misunderstanding between the Swiss's view of "the deal" (payment for results) and the bankers' (match the market). Alternatively, the Swiss may have claimed a misunderstanding as a way to cover their decision to lower bonuses after the junk-bond market collapsed and so bankers specializing in mergers and acquisitions became less valuable, reducing the present value of future profits from the relationship.

Some firms use formal and relational incentive contracts in combination. For example, recall that Lincoln Electric uses both piece rates and subjective bonuses. Baker, Gibbons and Murphy (1994) explore the simultaneous use of formal contracts based on objective performance measures p and relational contracts based on subjective assessments of total contribution y. The relational contract can reduce the distortionary incentives that would be created by the formal contract on its own, while the formal contract can reduce the size of the relational-contract bonus that the firm would need to offer if it used only a relational contract. Thus, the combination of formal and relational contracts can reduce distortion in the agent's incentives and reduce the firm's temptation to renege on a promised bonus.

All of the relational contracts discussed so far are incentive contracts, where current pay depends on current performance. But as noted above, subjective assessments of current performance may play a crucial role even if current compensation does not involve incentive pay of any kind. For example, the "fear of firing" models of Becker and Stigler (1974), Lazear (1979), and Shapiro and Stiglitz (1984) make continued employment contingent on a subjective performance assessment. MacLeod and Malcomson (1998) allow firms to use either or both kinds of relational contracts: subjective assessments of current performance influence both current compensation (performance pay) and continued employment (fear of firing). They prove that performance pay is more efficient than fear of firing when the cost of having a job vacant is low and qualified workers are in short supply.

Finally, subjective assessments can matter where workers have "career concerns," in finite-horizon settings as well as repeated games. In Holmström's (1982) model, for example, a worker's total contribution to firm value depends not only on the worker's effort but also on the worker's ability. Firms are initially uncertain about the worker's ability and so use observed performance to update their beliefs. Competition among prospective employers makes the worker's future wage depend on firms' updated beliefs about the worker's ability, so the worker has an incentive to perform well to influence these beliefs. Such career-concern models show that subjective assessments matter when future compensation depends on current performance. Of course, there is a related argument when future promotions depend on current performance, to which I now turn.

Skill Acquisition

In this section I shift the focus from incentives for various kinds of effort to incentives for skill acquisition. This shift makes performance evaluation trickier,

because the firm must now evaluate a worker's potential contribution to future firm value, rather than the worker's realized contribution to date. Accordingly, all the difficulties of objective performance measurement described above continue to apply, but probably with more force. Relational incentive contracts based on subjective performance assessments may again be attractive, but the shift from realized to potential performance probably makes it more difficult to find a subjective assessment that both the firm and the worker can observe. I therefore explore a third class of models, based on promotion rules rather than formal or relational incentive contracts. Promotion rules sometimes can induce the firm to deliver rewards based on subjective performance assessments. Furthermore, promotion rules can serve this function even in finite-horizon employment relationships, whereas the repeated-game models described above require the shadow of the future to influence the firm's current choices. I describe simple versions of Prendergast's (1993) model of promotions and Kahn and Huberman's (1988) model of an up-or-out rule; Gibbons and Waldman (1998) survey the broader literature on careers in organizations.

I start by defining an environment in which, if a court could verify a worker's potential contribution to future firm value, it would be simple to achieve efficient skill acquisition. This full-contracting world is the case analyzed by Becker (1962). But throughout this paper I have argued that it is often difficult for a court (or other third party) to determine a worker's realized contribution to firm value. And I have now suggested that it is even more difficult for a court to determine a worker's potential contribution to future firm value. This latter difficulty is especially pronounced regarding firm-specific (rather than general-purpose) human capital. That is, it might be merely hard to get outside evaluators to estimate the value of a worker's general-purpose human capital, but it may be nearly impossible for such outsiders to estimate the value of the worker's firm-specific human capital, since making such an estimate would require detailed knowledge about the firm's current and future operations and markets. In sum, just as the recent literature on objective performance measurement has rejected the strong but unremarked assumption in the classic agency model that the agent's total contribution is simply "output" that can easily be measured, the recent literature on the institutions that govern skill acquisition has rejected the parallel assumption in the classic human-capital model that the worker's potential contribution to future firm value is simply "skill" that again can easily be measured.

To be concrete, let y now denote the firm's assessment of the worker's potential contribution to future firm value, based on the worker's performance in an initial probationary period. This probationary period may be long—such as six years in many accounting and law partnerships and in academics. If the worker spends extra time during the probationary period learning about the firm's markets, competitors, technology, culture, and so on then the worker will have a high potential contribution for the future. More specifically, suppose the worker's potential contribution is x if the worker does not invest in such firm-specific human capital but is $x + v$ if the worker does invest. To make such an investment, the worker must

give up a certain amount of leisure time. Denote the opportunity cost of this time by c. Finally, suppose that the value of the investment (to the firm) exceeds the cost of the investment (to the worker); that is, $v > c$.

If a contract based on the worker's potential contribution could be enforced, it would be simple to achieve efficient skill acquisition in this setting. For example, the firm and the worker could sign a contract specifying that the firm will pay the worker a high wage if the worker achieves $y = x + v$ but a low wage if not. If the difference between these wages exceeds the worker's opportunity cost of investing, then the worker has an incentive to invest. Of course, the firm will want to induce such investment only if it receives a productivity increase that exceeds the wage increase. But since the value of this investment exceeds its cost, it is straightforward to find high and low wages that satisfy these two conditions. Finally, these high and low wages also need to make both parties willing to participate: for the worker, the high wage minus the opportunity cost must exceed the worker's best alternative opportunity; for the firm, the worker's productivity after investing must exceed the high wage. But this investment is not interesting unless the worker's productivity after investing minus the worker's cost of investing exceeds the worker's best alternative, so these two participation conditions are also simple to satisfy. In short, if a contract based on the worker's potential contribution could be enforced, efficient skill acquisition could be induced by contract, without recourse to institutions such as promotion rules.[4]

But suppose (here and for the remainder of this section) that contracts based on the worker's potential contribution cannot be enforced. Then we need an indirect way to induce the worker to invest *and* to induce the firm to reward such an investment. That is, non-contractible specific human capital creates a two-sided incentive problem; the worker is concerned that the firm cannot be trusted to reward investment properly, and the firm is concerned that the worker will not invest unless such rewards are anticipated. Prendergast (1993) shows that, under certain circumstances, the promise of promotion can solve this two-sided incentive problem, as follows.

Consider a simple model in which a firm has two jobs, easy and difficult. As before, suppose that the worker bears the opportunity cost c from investing in skills. Suppose also that such an investment improves productivity in both jobs, but that the productivity increase is greater in the difficult job. More specifically, suppose that: 1) an untrained worker is more productive in the easy job; 2) a trained worker is more productive in the difficult job; and 3) training is efficient, because the productivity difference between a trained worker in the difficult job and an untrained worker in the easy job exceeds the opportunity cost of training.

[4] Formally, suppose that w_h will be paid if $y = x + v$ but only w_l will be paid if $y = x$. Then the worker will choose to invest if $w_h - w_l > c$ and the firm will want to induce investment if $v > w_h - w_l$. Because $v > c$, it is simple to find wages that satisfy $v > w_h - w_l > c$. As for the participation constraints, let r denote the worker's best alternative opportunity. Then the worker requires $w_h - c > r$ and the firm $x + v - w_h > 0$. But this investment is not interesting unless $x + v - c > r$, so it is again simple to satisfy $x + v > w_h > r + c$.

Suppose that the firm can commit to paying a high wage in the difficult job and a low wage in the easy job. If the worker believes that investing in skills will yield a promotion to the difficult job, then the worker will invest if the difference between the high and low wages exceeds the opportunity cost of training. The firm, for its part, will choose to promote a trained worker if doing so is more profitable than leaving the trained worker in the easy job; that is, the productivity difference for a trained worker between the difficult and easy jobs must exceed the wage difference between the two jobs. Unfortunately, these two conditions may be incompatible, even if investment is efficient.[5] For example, suppose that an untrained worker produces 10 in the easy job, that a trained worker produces 20 in the easy job and 30 in the difficult job, and that the opportunity cost of training is 15. Then training is efficient ($30 - 10 > 15$) but we cannot find wages that simultaneously induce the worker to invest (wage difference greater than opportunity cost, 15) and induce the firm to promote a trained worker (wage difference smaller than productivity difference, $30 - 20$).

In short, an "up-or-stay" promotion rule (where the worker is either promoted or remains in the original job) creates a tension between needing a large enough wage gap to induce the worker to invest and keeping the wage gap small enough that the firm is willing to promote the worker after the worker has invested. If the value of the worker's specific capital in the two jobs is sufficiently different, then there exists a wage gap that meets both these constraints. But if the two jobs in question are really just two job titles sharing the same underlying technology, then an up-or-stay rule cannot simultaneously provide an incentive for the worker to invest and an incentive for the firm to reward investment.

I turn next to up-or-out rules—contracts specifying that after some fixed probationary period the firm must either pay the worker a high wage (w^*) or fire the worker. Up-or-out rules can induce the worker to invest in specific capital under circumstances where up-or-stay rules would fail, such as where the two jobs in question are really just two job titles sharing the same underlying technology. In keeping with this argument, up-or-out rules are often observed in accounting and law partnerships and in academics, where junior and senior jobs are often not hugely different.

Kahn and Huberman (1988) develop a model in which an up-or-out rule solves the two-sided incentive problem created by non-contractible specific human capital. As before, the investment is not interesting unless it is efficient. If the worker anticipates a promotion for investing, the worker will invest if the high wage (w^*) minus the opportunity cost exceeds the best alternative. The firm, for its part, will promote a worker who invested if the worker's productivity exceeds the high wage. In this simple model there is always a high wage that is both high enough to induce the worker to

[5] Formally, suppose the wages are w_e and w_d in the two jobs. The worker will invest if $w_d - c > w_e$, and the firm will promote a trained worker provided that $y_{dt} - w_d > y_{et} - w_e$, where y_{dt} and y_{et} are the productivities of a trained worker in the difficult and easy jobs respectively. We therefore require $y_{dt} - y_{et} > w_d - w_e > c$, but the assumption that training is efficient guarantees only that $y_{dt} - y_{en} > c$ (where y_{en} is the productivity of an untrained worker in the easy job), not that $y_{dt} - y_{et} > c$.

invest and yet low enough to induce the firm to promote a trained worker.[6] But this example of an up-or-out rule works too neatly. An up-or-out rule also may have big costs, even if it solves the two-sided incentive problem discussed here. Suppose, for example, that workers who make the appropriate investment could realize any one of several different levels of human capital. If some of the low realizations make the worker worth less than the high wage attached to the promotion then these workers will be fired, even though they undertook the appropriate investment. Firing these workers wastes their specific capital. Also, the prospect of possibly being fired means that workers have less incentive to invest in the first place. A similar problem arises if there is more than one up-or-out rung in a job ladder—those who clear the first hurdle but fail the second also represent wasted specific capital.

These examples identify a tradeoff between up-or-stay and up-or-out promotion rules. The up-or-stay rule induces the worker to invest in skills only if the two jobs are sufficiently different in how they utilize skill, whereas the up-or-out rule can induce investment even if training yields identical productivity increases in all jobs. On the other hand, the up-or-stay rule never wastes the acquired skills of those not promoted, whereas an up-or-out rule has this problem in many settings.

Incentive Contracts Between Versus Within Firms

All the models described above use the language of internal organization, as though they apply only to incentive contracts within firms, but none of them is limited to this context. For example, McMillan (1990) shows how several such models apply to supply relationships in the United States and Japan. Certainly the "get what you pay for" models of objective performance measurement could describe a supply transaction between two non-integrated firms. Similarly, the repeated-game models of subjective performance assessments could describe an ongoing "hand in glove" supply relationship. And even the promotion and up-or-out models could describe inducing a supplier to make specific investments by promising expanded responsibilities or by threatening termination.

So how can we tell if a particular model describes an incentive contract within or between firms, and does it matter? This is of course the Coase-Williamson problem of the boundary of the firm, applied to incentive contracts. A few papers address this issue and they show that it does indeed matter whether an incentive contract is within or between firms. The key accomplishment of all these models is that, for a given incentive problem, they derive the optimal incentive contract under both integration and non-integration (that is, within and between firms) and then compare the social surplus produced by each.

[6] Formally, in the notation introduced above, the investment is efficient if $x + v - c > r$. The worker will invest if $w^* - c > r$; the firm will promote a worker who has invested if $x + v - w^* > 0$. Thus, any value of w^* satisfying $x + v > w^* > r + c$ will induce the worker to invest and yet cause the firm to promote the worker.

Holmström and Milgrom (1991) develop a model of asset ownership, in the spirit of Grossman and Hart (1986) but with the addition of an incentive contract. The model is similar to their job-restrictions model described above: measured performance p reflects one action a_1 but another action a_2 changes the value v of an asset used in the production process. As in the job-restrictions model, these two actions compete for the agent's attention, in the sense that increasing the level of one increases the marginal cost of the other. The new feature of this asset-ownership model is that the agent can be either an employee or an independent contractor. An employee is defined as an agent who does not own the asset and so is paid only on measured performance ($w = s + bp$), whereas an independent contractor is defined as an agent who does own the asset and so receives both wages and any change in the asset's value ($w + v$). An agent who owns the asset then has an incentive to invest in it. But as in the job-restrictions model, if the agent is distracted by incentives for action a_2, it will require a larger bonus rate b to focus the agent's attention on action a_1. For different parameters—such as the extent to which achieving a high total contribution to firm value requires high levels of both actions simultaneously—either employment or independent contracting can be optimal. For a fixed set of parameters, the optimal contract ($w = s + bp$) is different for an employee than for an independent contractor. In particular, the optimal bonus rate b is lower for an employee, because the employee is not distracted by incentives to invest in the asset.

This asset-ownership model in Holmström and Milgrom (1991) is a model of incentive contracting that happens to focus on the boundary of the firm. A complementary paper by Holmström and Tirole (1991) can be seen as reversing the emphasis, asking how integration will affect not only incentive contracts but also sourcing decisions (for example, must a downstream division source only from the internal supplier upstream?) and product choices (will the internal supplier upstream develop products that are tailored for the downstream division or valued also by other downstream parties?). In the Holmström-Tirole model, non-integration means that there are two firms, each with its own principal and its own agent. In this case, as in Fershtman and Judd (1987), the problem with non-integration is a contracting externality; in designing its contract, neither firm takes account of the way the induced actions of its agent affect the other firm. Holmström and Tirole's innovation is not only to consider integration as a solution to this contracting externality, but also to note that integration brings costs on the sourcing and product-choice dimensions as well as benefits on the contracting dimension.

Both the Holmström and Milgrom (1991) and the Holmström and Tirole (1991) models are static. Beginning from a similar static model, Baker, Gibbons and Murphy (1997) develop a repeated-game model of subjective performance assessment that again allows for both integration and non-integration. Each period, an upstream party uses an asset to produce a good that could be used in a downstream party's production process. Ownership of the asset conveys ownership of the good. Thus, if the upstream party is an independent contractor then that party owns the good and so could sell it to a different downstream party, whereas if the upstream party is an employee then

the original downstream party already owns the good. The ownership of the good affects the parties' incentives to renege in the repeated game. If the downstream party owns the good then the model is identical to the version of Bull's (1987) model sketched above; in particular, the downstream party could keep the good without paying any of a promised bonus. But if the upstream party owns the good then two new considerations arise. First, the upstream party's threat to sell the good to an alternative downstream party limits the original downstream party's ability to renege on a promised bonus. But second, having the upstream party own the good creates an incentive for the upstream party to develop a good that has high value to alternative downstream parties, simply to improve the upstream party's bargaining position with the original downstream party. For different parameters—such as the extent to which actions that increase the value of the good to the original downstream party also increase its value to others—either employment or independent contracting can be optimal. And for a fixed set of parameters, higher-powered incentives are feasible under non-integration. That is, the optimal subjective bonus for producing a high-quality good can be higher for an independent contractor than for an employee, because if the upstream party owns the good then the downstream party cannot renege on the bonus yet keep the good.

Conclusion

I hope to have suggested why and how recent work on incentives in organizations has moved beyond the classic focus on the tradeoff between insurance and incentives. Risk remains an important issue but is now recognized as one issue among many. We have models of why firms get what they pay for, and of what they might do about it. We also have begun to understand the institutions that create incentives for skill acquisition rather than simply for effort. Finally, we have begun to see why and how agency theory should become better integrated with Coase-Williamson literature on organizational economics more generally.

The central lesson for incentive contracting from the organizational economics literature is the old but important notion of fit (or complementarity, as it is now called). We have seen, for example, how it may be useful to impose job restrictions to reduce an agent's distractions, and that reducing the agent's outside interests (such as through changing asset ownership) can play a similar role. Once such distractions are reduced, the optimal incentive contract may well have a low bonus rate. In this sense, job restrictions, asset ownership, and low-powered incentives may be complementary. This systems perspective on incentive contracting is best articulated by Holmström and Milgrom (1994, p. 989), who reach an important conclusion: "The use of low-powered incentives within the firm, although sometimes lamented as one of the major disadvantages of internal organization, is also an important vehicle for inspiring cooperation and coordination."

The systems perspective also has broader implications: not only should complementary instruments be used in incentive contracting, but the firm's strategy

towards incentive contracting should complement its strategies towards human resource management, manufacturing, product development, and competition. For example, Ichniowski, Shaw and Prennushi (1997) use data painstakingly collected from 36 steel mills on incentive pay, recruiting and selection, teamwork, employment security, flexible job assignment, skills training, communication, and labor relations. They find that mills using bundles of complementary practices are more productive. Milgrom and Roberts (1995) and Freeman and Kleiner (1998) offer similar analyses of Lincoln Electric and a major U.S. shoe manufacturer, respectively. Much of the promising new work in this area comes from field researchers, especially in industrial relations and human resource management.

While I have focused on incentives for effort and skill acquisition, there are of course many other incentives in organizations, some of which might be labeled "politics." Unlike the classic agency model, most hierarchical relationships involve non-contractible decisions by the superior that affect the life of the subordinate. Bull's (1987) model fits this abstract description; Tirole (1986) and Milgrom and Roberts (1988) draw on the rich sociological literature on organizational politics to produce appealing models of collusion and influence activities in hierarchical relationships. The Tirole and Milgrom-Roberts models only begin to tap the large literature in organizational sociology that is just waiting to be explored by economists; see Gibbons (1998) for suggestions about how the two disciplines might complement each other in this domain.

Finally, economists are not the only ones thinking about incentives, even incentives for effort. Some economists have been reading these other literatures. Kreps (1997), for example, re-examines the old question of intrinsic versus extrinsic motivation; for influential work on this question from psychology, see Staw (1977). As another example, Rotemberg (1994) models an employment relation that is also a social relation; for an influential survey in this spirit, see Baron (1988). One simple possibility is that economic models that ignore social psychology are incomplete (but perhaps still useful) descriptions of incentives in organizations. A more troubling possibility is that management practices based on economic models may dampen (or even destroy) non-economic realities such as intrinsic motivation and social relations. Field experiments on this issue would be especially useful.

In sum, I think the economics literature on incentives in organizations has made important progress in the 1990s, opening up several new areas of inquiry beyond the classic focus on the tradeoff between insurance and incentives. But I think that much of the best economics on this subject is still to come, and that it will exhibit stronger connections both to the broader literature on organizational economics and to other disciplines that study organizations.

■ *I thank Brad De Long, Oliver Hart, Alan Krueger, and especially Timothy Taylor for helpful comments and George Baker, Bengt Holmström, Ed Lazear, Kevin J. Murphy, Canice Prendergast, and Mike Waldman for years of discussions.*

References

Aggarwal, Rajesh, and Andrew Samwick, "The Other Side of the Tradeoff: The Impact of Risk on Executive Compensation," *Journal of Political Economy,* forthcoming.

Alston, Lee, and Robert Higgs, "Contractual Mix in Southern Agriculture since the Civil War: Facts, Hypotheses, and Test," *Journal of Economic History,* 1982, *42,* 327–53.

Anderson, Kathryn, Richard Burkhauser, and Jennie Raymond, "The Effect of Creaming on Placement Rates under the Job Training Partnership Act," *Industrial and Labor Relations Review,* 1993, *46,* 613–24.

Baker, George, "Incentive Contracts and Performance Measurement," *Journal of Political Economy,* 1992, *100,* 598–614.

Baker, George, Robert Gibbons, and Kevin J. Murphy, "Subjective Performance Measures in Optimal Incentive Contracts," *Quarterly Journal of Economics,* 1994, *109,* 1125–56.

Baker, George, Robert Gibbons, and Kevin J. Murphy, "Relational Contracts and the Theory of the Firm." Unpublished manuscript, MIT's Sloan School, 1997.

Baron, James, "The Employment Relation as a Social Relation," *Journal of the Japanese and International Economies,* 1988, *2,* 492–525.

Becker, Gary, "Investment in Human Capital: A Theoretical Analysis," *Journal of Political Economy,* 1962, *70,* 9–49.

Becker, Gary, and George Stigler, "Law Enforcement, Malfeasance and Compensation of Enforcers," *Journal of Legal Studies,* 1974, *3,* 1–18.

Brown, Keith, W. Harlow, and Laura Starks, "Of Tournaments and Temptations: An Analysis of Managerial Incentives in the Mutual Fund Industry," *Journal of Finance,* 1996, *51,* 85–110.

Bull, Clive, "The Existence of Self-Enforcing Implicit Contracts," *Quarterly Journal of Economics,* 1987, *102,* 147–59.

Chevalier, Judith, and Glen Ellison, "Risk Taking by Mutual Funds as a Response to Incentives," *Journal of Political Economy,* 1997, *105,* 1167–1200.

Coase, Ronald, "The Nature of the Firm," *Economica,* 1937, *4,* 386–405.

Cockburn, Ian, Rebecca Henderson, and Scott Stern, "Balancing Research and Production: Internal capital markets and promotion policies as incentive instruments." Unpublished manuscript. MIT's Sloan School, 1998.

Compte, Olivier, "Communication in Repeated Games with Imperfect Private Monitoring," *Econometrica,* 1998, *66,* 597–626.

Cragg, Michael, "Performance Incentives in the Public Sector: Evidence from the Job Training Partnership Act," *Journal of Law, Economics, and Organization,* 1997, *13,* 147–68.

Eccles, Robert, and Dwight Crane, *Doing Deals: Investment Banks at Work.* Boston: Harvard Business School Press, 1988.

Fast, Norman, and Norman Berg, "The Lincoln Electric Company." Harvard Business School Case #376–028, 1975.

Fershtman, Chaim, and Kenneth Judd, "Equilibrium Incentives in Oligopoly," *American Economic Review,* 1987, *77,* 927–40.

Freeman, Richard, and Morris Kleiner, "The Last American Shoe Manufacturers: Changing the Method of Pay to Survive Foreign Competition." Unpublished manuscript, Harvard University, 1998.

Gibbons, Robert, "Taking Coase Seriously." *Administrative Science Quarterly,* forthcoming, 1998.

Gibbons, Robert, and Michael Waldman, "Careers in Organizations: Theory and Evidence." In Ashenfelter, O., and D. Card, eds., *Handbook of Labor Economics, Volumes III and IV.* New York: North Holland, 1998.

Grossman, Sanford, and Oliver Hart, "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Ownership," *Journal of Political Economy,* 1986, *94,* 481–510.

Hayes, Rachel, and Scott Schaefer, "Implicit Contracts and the Explanatory Power of Top Executive Compensation for Future Performance." Unpublished manuscript, Kellogg Graduate School of Management, Northwestern University, 1997.

Higgs, Robert, "Race, Tenure, and Resource Allocation in Southern Agriculture, 1910," *Journal of Economic History,* 1973, *33,* 149–69.

Holmström, Bengt, "Managerial Incentive Problems—A Dynamic Perspective." In *Essays in Economics and Management in Honor of Lars Wahlbeck.* Helsinki: Swedish School of Economics, 1982.

Holmström, Bengt, and Paul Milgrom, "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design," *Journal of Law, Economics, and Organization,* 1991, *7,* 24–52.

Holmström, Bengt, and Paul Milgrom, "The Firm as an Incentive System," *American Economic Review,* 1994, *84,* 972–91.

Holmström, Bengt, and Jean Tirole, "Transfer

Pricing and Organizational Form," *Journal of Law, Economics, and Organization*, 1991, *7*, 201–28.

Ichniowski, Casey, Kathryn Shaw, and Giovanna Prennushi, "The Effects of Human Resource Management Practices on Productivity: A Study of Steel Finishing Lines," *American Economic Review*, 1997, *87*, 291–313.

Kahn, Charles, and Gur Huberman, "Two-sided Uncertainty and 'Up-or-Out' Contracts," *Journal of Labor Economics*, 1988, *6*, 423–44.

Kandori, Michihiro, and Hitoshi Matsushima, "Private Observation, Communication and Collusion," *Econometrica*, 1998, *66*, 627–52.

Kerr, Steven, "On the Folly of Rewarding A, While Hoping for B," *Academy of Management Journal*, 1975, *18*, 769–83.

Kole, Stacey, "The Complexity of Compensation Contracts," *Journal of Financial Economics*, 1997, *43*, 79–104.

Kreps, David, "Intrinsic Motivation and Extrinsic Incentives," *American Economic Review*, 1997, *87*, 359–64.

Lazear, Edward, "Why Is There Mandatory Retirement?" *Journal of Political Economy*, 1979, *87*, 1261–84.

Lazear, Edward, "Pay Equality and Industrial Politics," *Journal of Political Economy*, 1989, *97*, 561–80.

Lazear, Edward, "Performance Pay and Productivity." Unpublished manuscript, Stanford University, 1997.

Lazear, Edward, and Sherwin Rosen, "Rank-Order Tournaments as Optimum Labor Contracts," *Journal of Political Economy*, 1981, *89*, 841–64.

MacLeod, Bentley, and James Malcomson, "Motivation and Markets," *American Economic Review*, 1998, *88*, 388–411.

Malcomson, James, "Incentive Contracts in Labor Markets." In Ashenfelter, O., and D. Card, eds., *Handbook of Labor Economics, Volumes III and IV*. New York: North Holland, 1998.

McMillan, John, "Managing Suppliers: Incentive Systems in Japanese and U.S. Industry," *California Management Review*, 1990, *32*, 38–55.

Milgrom, Paul, and John Roberts, "An Economic Approach to Influence Activities in Organizations," *American Journal of Sociology*, 1988, *94*, S154–79.

Milgrom, Paul, and John Roberts, "Complementarities and Fit: Strategy, Structure, and Organizational Change in Manufacturing," *Journal of Accounting and Economics*, 1995, *19*, 179–208.

Murphy, Kevin J., "Executive Compensation." In Ashenfelter, O., and D. Card, eds., *Handbook of Labor Economics, Volumes III and IV*. New York: North Holland, 1998.

Prendergast, Canice, "The Role of Promotion in Inducing Specific Human Capital Acquisition," *Quarterly Journal of Economics*, 1993, *108*, 523–34.

Prendergast, Canice, "The Provision of Incentives in Firms," *Journal of Economic Literature*, forthcoming.

Rotemberg, Julio, "Human Relations in the Workplace," *Journal of Political Economy*, 1994, *102*, 684–717.

Shapiro, Carl, and Joseph Stiglitz, "Equilibrium Unemployment as a Discipline Device," *American Economic Review*, 1984, *74*, 433–44.

Staw, Barry, "Motivation in Organizations: Towards Synthesis and Redirection." In B. Staw and G. Salancik, eds., *New Directions in Organizational Behavior*. Chicago: St. Clair Press, 1977.

Stewart, James, "Taking The Dare," *The New Yorker*, July 26, 1993, 34–39.

Tirole, Jean, "Hierarchies and Bureaucracies: On the Role of Collusion in Organizations," *Journal of Law, Economics, and Organization*, 1986, *2*, 181–214.

Williamson, Oliver, *Markets and Hierarchies*. New York: The Free Press, 1975.

Williamson, Oliver, *The Economic Institutions of Capitalism*. New York: The Free Press, 1985.