Low-Powered Incentives and the Motivation of Critical Workers

Joseph A. Ritter*
Lowell J. Taylor**

Current version: June 2004

JEL classification: J33

* Hubert H. Humphrey Institute of Public Affairs, University of Minnesota, Minneapolis, MN 55455

** H. John Heinz III School of Public Policy and Management, Carnegie Mellon University, Pittsburgh, PA 15213

Abstract

The paper revolves around the interaction of three ideas: (1) Because of the nature of some workplace technologies, "critical workers" can have a disproportionately negative effect on a firm's value added. (2) In many firms managers have little verifiable information about performance. (3) There are unobservable productivity differences among individuals. In this context we study the properties of low-powered incentives, in which compensation is not directly tied to performance and all but clearly unsatisfactory employees are retained. Low-powered incentives are compared to a potentially viable high-powered alternative, a rank-order tournament with pre-announced prizes. Low-powered incentives are found to be effective in motivating critical workers, while tournaments can be dysfunctional in that regard. The latter result stems from an "uneven playing field effect" in the tournament: By encouraging high-productivity workers, tournaments disadvantage and discourage low-productivity workers.

If a second violinist in a symphony orchestra plays out of tune, the Mahler is ruined. Economic intuition suggests that strong and direct incentives are appropriate here, yet symphony orchestras typically use only low-powered incentives—compensation is not directly tied to performance and all but clearly unsatisfactory employees are retained. This paper suggests one reason why organizations with critical employees often use primarily low-powered incentives.

The paper revolves around the interaction of three ideas: (1) Certain workers can have a disproportionately negative effect on a firm's value added. We call these *critical* workers. (2) In many firms managers have little or no contractible, or even common-knowledge, information about performance. Instead they rely on private impressions, either well or badly informed. (3) There are unobservable productivity differences among individuals. We discuss these observations in turn.

The idea that technology makes certain workers critical draws heavily on Kremer's (1993) theory of o-ring production. Critical workers are an important feature of some specialized settings, such as symphony orchestras, but workers can be be critical in more mundane workplaces as well: Sub-par work performance in one step of a manufacturing process can lead to costly recalls or destroy a product's reputation. Poor hygiene by a cook's assistant can lead to food poisoning and expensive lawsuits. In cases like these, minimum performance is most important to the firm. A technology like berry picking, by contrast, makes average performance more important.

A pivotal feature of any model of an agency problem is the nature of the information available to the players. We posit an environment that is reminiscent of the symphony orchestra, and which we believe to be a common one. We assume that managers do observe worker performance and form opinions about this performance, but these impressions are private and imperfect; the only contractible information is the worker's presence and the level of compensation. Thus any variation from fixed compensation must be at the discretion of the firm, not a feature of a contract. This greatly restricts choice of high-powered incentives; our attention centers on rank-order tournaments with pre-announced, verifiable prizes (Lazear and Rosen, 1981; Malcomson, 1984).

We establish a number of baseline results about how low-powered incentives function when workers are homogeneous. Then, following Baker, Gibbons, and Murphy (1994), we allow workers' productivity to vary unobservably, leading to imperfectly observable performance variation. Consequently, it is more costly to motivate a low-productivity worker to reach any given performance level than to motivate a high-productivity worker, but the former case is disproportionately important for a critical worker.

Our main results contrast low- and high-powered incentives as workers become, in a precise sense, "more critical." If workers become more critical, firms that use low-powered incentives can adjust their wages and retention policy to improve performance of low-productivity workers relative to high-productivity workers. Our key results about tournaments, however, are almost the reverse: High-productivity workers respond more vigorously than low-productivity workers to a larger prize; average performance increases more rapidly than minimum performance. In some circumstances, increasing the size of the prize can even de-motivate low-productivity workers. Theese results stem from an "uneven playing field effect" in the tournament, which offsets the direct motivation of competition: By encouraging high-productivity workers, increasing the size of the prize progressively disadvantages and discourages low-productivity workers.

The relationship between low-powered incentives and tournaments in our model is somewhat subtle. Low-powered incentives address an agency problem by giving the job positive value. Our particular version of low-powered incentives is similar to efficiency wage models in that the job's value comes from a wage premium, which makes resolution of the agency problem costly to the firm.[1] The theoretical appeal of rank-order tournaments (and other high-powered incentives) is that they can motivate workers without lowering profits by changing how a given level of compensation is distributed. We show, in fact, that a firm can always increase profits in our model by supplementing low-powered incentives with a small tournament.[2] It does not follow, however, that the firm will want to *replace* low-powered incentives with prizes based on internal competition. Besides the possibility that the size of prizes in a zero-cost tournament may be limited by various considerations, tipping compensation more heavily toward the use of prizes can be de-motivating to low-productivity workers, a problem that is clearly germane for critical workers.

The paper proceeds as follows. Section I poses the incentive problem and a model of low-powered incentives. Section II establishes some baseline results about low-powered incentives with homogeneous workers. Section III introduces unobservable heterogeneity, gives a precise meaning to "more critical," and shows how low-powered incentives adapt to this aspect of the production technology. Section IV generalizes the model, allowing compensation policies that mix low-powered incentives and a rank-order tournament. Section V concludes.

I. The Model

The performance levels of N workers determine the output of a profit-maximizing firm:

$$Y = G(p^1, p^2, \ldots, p^N)$$

[1] This is not a characteristic of low-powered incentives, per se. Lazear's (1979) model of upward-sloping earnings profiles does not tie compensation directly to performance, but at least partly resolves the agency problem without lowering profits.

[2] We do not address problems with implementing tournaments that others have raised: Prendergast and Topel (1996) have argued that influence activity discourages the use of this form of incentive pay. Also, Lazear (1989) has noted that when cooperation among workers is important, salary compression (that is, movement toward lower-powered incentives) is optimal.

with $\partial G/\partial p^i > 0$. We assume that Y provides no usable information about p_i. Formally, our assumption about the information content of Y means that we simply do not consider compensation schemes that condition pay on Y. Informally, we interpret this assumption to mean that either there are no variables related to a worker's contribution to firm value on which either an explicit or implicit contract can be based or, in any event, that the firm should not use them because they would provide dysfunctional incentives for any of the various reasons discussed in the literature.[3] Initially, we rule out strategic interactions among workers—they do not condition their behavior on the behavior of other workers. (In Section IV, we introduce internal competition.)

Since workers appear identical to the firm, we assume they are treated alike, and we work with the function

$$g(p^i) = G(p^i, p^{(-i)}),$$

where $p^{(-i)}$ denotes the performance levels of workers other than i. We assume that $g(p)$ has continuous first and second derivatives and is concave. Also $g'(\infty) = 0$.

We abstract from life-cycle effects by assuming that workers are infinitely-lived (or, equivalently, face a Poisson probability of death). In each period workers choose a scalar performance level p, known only to the employee. A worker's utility in each period is the wage less her performance level: $w - p$. (In Section III we induce productivity differentials by adding variation in the disutility ("effort") required to achieve a given performance level.)

Employees are paid for one unit of labor per period. They have a discount rate ρ and discount factor $\beta = 1/(1+\rho)$.[4] The expected value of a worker's alternative

[3] Gibbons (1998) includes a survey of these reasons. Prendergast (1999) summarizes: "Perhaps the most striking aspect of observed contracts is that the Informativeness Principle, i.e., that all factors correlated with performance should be included in a compensation contract, seems to be *violated* in many occupations."

[4] A constant exogenous probability of separation would enter the model in a fashion almost identical to β.

activities is V^a. Since we model only the behavior of a single firm, there is no reason to be more specific about these alternate activities. Hiring and termination are costless to the firm. Each employee is supervised by a single manager. We assume the manager behaves in the firm's interests (though this is a subject of considerable independent interest).

The firm's information about the performance level of a particular worker is minimal: The manager receives a noisy signal of the worker's performance,

$$x = p + \epsilon.$$

We assume that ϵ has zero mean, and that its distribution has a differentiable, symmetric single-peaked density f. We denote its cumulative density by F. We interpret x as a manager's impression or opinion of how well the worker is performing. The realization of x is private information and, therefore, cannot reasonably form the explicit basis for any compensation policy.[5]

We base our model on subjective assessments because many important components of performance—politeness, enthusiasm, or cooperativeness—are difficult to define explicitly, or even to articulate, but can nonetheless be assessed by an observant manager. Thus, a manager can often reasonably determine whether a worker is performing well in a multi-tasking environment or where output is intangible.[6]

Although x is private information, the density f summarizes characteristics of the manager and production environment that are common knowledge. For example, a clueless manager has a relatively diffuse distribution for ϵ. Similarly, differences among firms' production processes also generate cross-sectional variation in

[5] Baker, Gibbons, and Murphy (1994) model subjective assessments as common knowledge rather than private information.

[6] With respect to multi-tasking, MacLeod and Parent (1999) argue that even where explicit performance measurements are available for each task, the combinatorial demands of optimally balancing them in an explicit agreement quickly become overwhelming.

f. If workers are physically separated from managers, for example, managers' impressions would tend to be more diffuse. The span of a manager's control, explicitly modeled by Mehta (1998), would also affect f.

Simpler models of low-powered incentives have been studied under the headings of efficiency wages and deferred compensation (for example, Shapiro and Stiglitz, 1984, and Lazear, 1979). The model just outlined is distinctive in two ways. First, employee's choice of performance level is not restricted. In most earlier models "effort" is taken to be a discrete variable with only one economically viable value (i.e., agency problems must be fully resolved). Second, the "monitoring" technology, though exogenously determined, is continuous, while previous models generally assume an exogenous binary signal.

II. Employment Policy with Identical Workers

Since workers do not interact strategically with one another, the model is a repeated game between the firm and a single worker with the following order of play in each round: (1) The firm offers a wage w. (2) The worker responds with a performance level $p \geq 0$. (3) Nature plays x using the distribution F. (4) The firm pays w. (5) The firm decides whether to retain the worker or end the game. We focus on repeated play of the Bayesian Nash equilibrium of this stage game in which the worker is retained if and only if x exceeds an endogenous threshold $\bar{x}$. We assume that $\bar{x}$ is common knowledge. (Workers could infer $\bar{x}$ by observing the frequency of terminations.)

A. The Employee's Problem

Let $\hat{p}$ be a worker's best response to an employment policy $\psi = \{w, \bar{x}\}$ that satisfies the worker's participation constraint: $V(\hat{p};\psi) \geq V^a$, where $V(\hat{p};\psi)$ is the maximum utility that can be achieved by accepting ψ. (To simplify notation, we forthwith suppress ψ as an argument in V.) The lifetime utility of an employee who chooses performance level p today and reverts to $\hat{p}$ tomorrow is given by

$$V(p) = w - p + \beta\big[F(\bar{x} - p)V^a + \big(1 - F(\bar{x} - p)\big)V(\hat{p})\big]. \tag{1}$$

The employee maximizes $V(p)$ by choosing $p \geq 0$. The worker's best response is well-behaved:

Proposition 1: *An employment policy $\psi = \{w, \bar{x}\}$ that satisfies the participation constraint implies a unique best response $\hat{p}$. Further, $\hat{p}$ exceeds the termination threshold $\bar{x}$ if $V(\hat{p}) > V^a$.*

Proof: First note that $V(\hat{p}) = V^a$ implies that $\hat{p} = 0$. Assume now that $V(\hat{p}) > V^a$. Then $V(p)$ is continuous in p, and $\lim_{p\to\infty} V(p) = -\infty$, so $V(p)$ must have a maximum for $p > 0$.[7] Thus for today's best action to be $\hat{p}$, we must have $V'(\hat{p}) = 0$, which reduces to

$$\beta f(\bar{x} - \hat{p})[V(\hat{p}) - V^a] = 1. \tag{2}$$

We also have

$$V''(p) = \beta f'(\bar{x} - p)[V^a - V(\hat{p})],$$

which is negative long as $f'(\bar{x} - p) > 0$. Thus $V(p)$ is concave where $\bar{x} - p < 0$. Therefore, any solution to equation (2) is a maximum if $\bar{x} - \hat{p} < 0$. To see that the solution is unique, note that two maxima would bracket a minimum, where $V''(p) > 0$, implying that $f'(\bar{x} - p)$ switches sign twice. But this contradicts the assumption that $f(\epsilon)$ is single-peaked. ∎

According to Proposition 1, workers are terminated only when their perceived performance is strictly less than their actual performance. The economic intuition is that the worker, recognizing the imperfection of the signal, establishes a buffer $(\hat{p} - \bar{x})$ between her actual performance the perceived performance level that results in

[7] An arbitrary ψ might imply the corner solution $p = 0$. But obviously we have assumed the firm wants to address its agency problem rather than ignore it. Therefore we ignore this detail.

termination. An outsider, somehow able to observe both $\hat{p}$ and $\bar{x}$, might misinterpret this outcome in two ways. The observer might conclude that managers tolerate substandard performance, firing workers only when they see performance well below the norm. Alternatively, the observer might interpret $\hat{p}-\bar{x}$ as a "gift" of performance exceeding some standard or required level, but in fact there is no reciprocity in this model.

Evaluating equation (1) at $p = \hat{p}$ and solving for $V(\hat{p})$, then substituting into equation (2) produces

$$w = \hat{p} + (1-\beta)V^a + \frac{\rho + F(\bar{x}-\hat{p})}{f(\bar{x}-\hat{p})}, \tag{3}$$

which implicitly defines the worker's best response $\hat{p}(\psi)$. Since $w - \hat{p}$ is the flow of utility from the job and $(1-\beta)V^a$ is the flow value of alternative activities, equation (3) guarantees that any ψ for which $\hat{p} > 0$ also satisfies the participation constraint.

In contrast to models that assume 0/1 effort decisions, here $\hat{p}$ is continuous in ψ. The former models create artificial nonconvexities so that slight deviations from optimal compensation can result in a discrete jump to $p = 0$.

B. Profit Maximization

The firm's objective is to maximize profits, which it must do subject to the constraint imposed by the worker's best response function $\hat{p}(\psi)$:

$$\max_{\psi} g(\hat{p}(\psi)) - w.$$

We assume positive profits are possible. Profits are bounded since we have assumed $g'(\infty) = 0$. This problem is easier to handle if we observe that equation (3) has a dual interpretation as specifying the minimum w required to induce performance $\hat{p}$ for a given $\bar{x}$.[8] We thus write equation (3) as $w = w(\hat{p}, \bar{x})$ and use it to formulate

[8] That is,

$$w(\hat{p}, \bar{x}) = \min_{w}\{w \text{ subject to } (3)\}.$$

an equivalent profit maximization:

$$\max_{\bar{x},\hat{p}} g(\hat{p}) - w(\hat{p}, \bar{x}). \tag{4}$$

We will continue to use $\hat{p}$ to denote the worker's best response, while using p^* to denote the performance induced by the firm's optimal choice of ψ.

We now turn to the question of how efficiency and distribution of surplus are affected by the agency problem. With regard to efficiency, the following result follows easily from the modified profit-maximization problem (4):

Proposition 2: *The optimal employment policy ψ^* induces the first-best performance level:*

$$g'(p^*) = 1.$$

Proof: The first-order conditions for (4) are

$$\begin{aligned} 0 &= g'(p^*) + \frac{\partial w(p^*, \bar{x}^*)}{\partial p^*} \\ 0 &= \frac{\partial w(p^*, \bar{x}^*)}{\partial \bar{x}}. \end{aligned}$$

Since both $\hat{p}$ and $\bar{x}$ are parameters in the minimization implied by the dual interpretation of (3), we can apply the envelope theorem to $w(\hat{p}, \bar{x})$ (differentiate the right-hand side of (3)). By inspection,

$$\frac{\partial w(\hat{p}, \bar{x})}{\partial \hat{p}} = 1 - \frac{\partial w(\hat{p}, \bar{x})}{\partial \bar{x}},$$

and thus $g'(p^*) = 1$. ∎

Proposition 3 describes the distribution of surplus and retention policy.

Proposition 3: *Suppose that the density f depends on a parameter σ, with σ^2 linearly related to the variance, and that f can be written*

$$f(\epsilon; \sigma) = \frac{1}{\sigma} h\left(\frac{\epsilon}{\sigma}\right) \tag{5}$$

where h does not depend on σ except via ϵ/σ. Then:

i. *The probability of retention does not depend on σ.*

ii. *Compensation is linearly increasing in σ:*

$$w^* = p^* + (1-\beta)V^a + \frac{h(b^*)}{h'(b^*)}\,\sigma. \tag{6}$$

Proof: Make the change of variables $b = y/\sigma$, so that

$$f(\sigma b) = \frac{h(b)}{\sigma}.$$

It is easy to show that

$$f'(\sigma b) = \frac{h'(b)}{\sigma^2}.$$

From profit maximization we have

$$0 = \frac{\partial w(p^*, \bar{x}^*)}{\partial \bar{x}} = 1 - \frac{\big[\rho + F(\bar{x}^* - p^*)\big] f'(\bar{x}^* - p^*)}{f(\bar{x}^* - p^*)^2}$$

Solving for $F(\bar{x} - \hat{p})$ and substituting for f and f' makes it clear that the retention probability is independent of σ:

$$F(\bar{x}^* - p^*) = \frac{h(b^*)^2}{h'(b^*)} - \rho$$

Substitution into equation (3) produces equation (6). ∎

Condition (5) is met by a wide range of distributions including normal, logistic, and nonstandard t variables, as well any mixture of them (though some mixture distributions violate the assumption that f is single-peaked).

Combined with Proposition 2, this result demonstrates that in the simple setting we model, bad managers or difficult environments make it costly to achieve efficient performance. One might expect that in an increasingly noisy environment the firm would respond by trading off lower expected performance for lower compensation. Instead, an increase in σ only moves the worker farther away from her participation constraint (while $\sigma = 0$ makes the participation constraint bind).

III. Critical Workers, Heterogeneity, and Low-Powered Incentives

Whether a worker is critical makes no difference in the baseline model described in the previous section. In this section we show that when workers' productivity varies unobservably, the form of the production technology influences the design of optimal low-powered incentives.

Following Baker, Gibbons, and Murphy (1994), the form of heterogeneity we introduce is transitory variation in the work environment. If workers are better positioned than supervisors to observe certain stochastic elements in the production process, then the motivational effect of any incentive scheme that relies on the supervisor's assessment (e.g., low-powered incentives) varies unobservably with circumstances. Employers with different production technologies therefore differ in their tolerance for heterogeneous performance. Thus they must concern themselves with the effect of their employment policy on more than one "type" of worker.

Specifically, suppose that the effort required to achieve a given level of performance is random. The variation can come from either the supply side (the worker is not feeling healthy) or the demand side (unusual problems must be solved). The disutility incurred in supplying performance p is ηp. The random variable η is binary with $\eta_1 > \eta_2$, $\Pr\{\eta = \eta_1\} = \theta$, and $\mathrm{E}\{\eta\} = 1$. Each realization of η is independent of past realizations and those of other workers. We assume that the realization of η is private information, known only to the worker. We continue to assume that $x = p + \epsilon$ and output is $g(p)$.

As in Section II, the problem we solve is how to design the employment policy ψ that should be offered to worker i holding other workers' performance choices fixed. We first consider the worker's best response to employment conditions ψ. The expected lifetime utility of a worker whose current type is j is

$$V(p_j) = w - \eta_j p_j + F(\bar{x} - p_j)\beta V^a + \big(1 - F(\bar{x} - p_j)\big)\beta \mathrm{E}_\eta\{V\}, \tag{7}$$

where $\mathrm{E}_\eta\{V\} = \theta V(\hat{p}_1) + (1-\theta)V(\hat{p}_2)$. The first-order conditions for maximizing V with respect to p_j are

$$\eta_j = \beta[\mathrm{E}_\eta\{V\} - V^a]f(\bar{x} - \hat{p}_j), \quad j = 1, 2. \tag{8}$$

The structure of the employee's maximization is much the same as in the simpler model, and thus Proposition 1 applies to both $\hat{p}_1$ and $\hat{p}_2$. Note that equations (8) imply

$$\frac{f(\bar{x} - \hat{p}_1)}{f(\bar{x} - \hat{p}_2)} = \frac{\eta_1}{\eta_2}. \tag{9}$$

Since second-order conditions imply $f'(\bar{x} - p_j) > 0$, equation (9) implies that $\hat{p}_1 < \hat{p}_2$.

Taking expectations of both sides of equation (7) with respect to η gives an expression for $\mathrm{E}_\eta\{V\}$ that can be used in equation (8). After rearranging

$$w = \theta\eta_1\hat{p}_1 + (1-\theta)\eta_2\hat{p}_2 + (1-\beta)V^a + Q(w, \bar{x})\left(\frac{\eta_j}{f(\bar{x} - \hat{p}_j)}\right), \quad j = 1, 2 \tag{10}$$

where $Q(w, \bar{x}) = \rho + \theta F(\bar{x} - \hat{p}_1) + (1-\theta)F(\bar{x} - \hat{p}_2)$. Equations (10) implicitly define the worker's best responses $\hat{p}_1$ and $\hat{p}_2$ to employment policy ψ.

We list several useful characteristics of the best responses in Proposition 4, which we prove in the appendix.

Proposition 4: *Suppose $f(\bar{x} - p_2)/f(\bar{x} - p_1)$ is increasing in $\bar{x}$ (that is, f possesses the monotone likelihood ratio property), then the worker's best responses have the following properties:*

$$\frac{\partial \hat{p}_j}{\partial w} > 0 \qquad \textit{and} \qquad \frac{\partial \hat{p}_j}{\partial w} + \frac{\partial \hat{p}_j}{\partial \bar{x}} = 1.$$

Additionally, at ψ^*,

$$\frac{\partial \hat{p}_1^*}{\partial w} > 1, \qquad \frac{\partial \hat{p}_1^*}{\partial \bar{x}} < 0, \qquad \textit{and} \qquad \frac{\partial \hat{p}_2^*}{\partial \bar{x}} > 0.$$

Although it is not surprising that the derivatives of $\hat{p}_j$ with respect to the wage are both positive, it is important that an increase in wage has a larger marginal effect on performance of the low-productivity workers ($\hat{p}_1$) than on the high-productivity workers ($\hat{p}_2$). The derivatives with respect to dismissal threshold $\bar{x}$ are more interesting. With *homogenous* performance, the firm, for any given w, adjusts $\bar{x}$ so as to maximize performance: $\partial\hat{p}/\partial\bar{x} = 0$. With *heterogeneous* performance the firm chooses a compromise $\bar{x}$, which is "too high" from the perspective of achieving the highest possible $\hat{p}_1$ and "too low" from the perspective of achieving maximum $\hat{p}_2$. As a consequence, for a given wage, the firm can alter the variability in performance by adjusting retention policies; if the firm reduces $\bar{x}$ (i.e., retains a higher fraction of the work force), $\hat{p}_1$ increases while $\hat{p}_2$ declines, reducing variation in worker performance.

The claims of Proposition 4 suggest that a high-wage, low-turnover policy might be optimal for a firm that is concerned with maintaining a relatively high minimal level of performance (i.e., motivating p_1). Thus an orchestra might use a high-retention policy, using dismissals only for extremely poor performance.[9] We proceed by investigating this conjecture.

The contribution of a single worker is now $g(\hat{p}_1)$ with probability θ and $g(\hat{p}_2)$ with probability $1 - \theta$.[10] It is relatively easy to establish (using the first part of Proposition 4) that when ψ^* maximizes $\theta g(\hat{p}_1) + (1-\theta)g(\hat{p}_2) - w$ subject to equations (10),

$$\theta g'(p_1^*) + (1-\theta)g'(p_2^*) = 1. \tag{11}$$

Since $\hat{p}_1 < \hat{p}_2$ and, by assumption, $g'(1) = 1$, it follows from equation (11) that $p_1^* < 1 < p_2^*$.

[9] In fact, elite symphony orchestras take considerable care to hire strong musicians, but then adopt high-retention policies for these musicians. Though we do not model it here, there is an additional reason why the the low-powered incentive structure will be effective in dealing with heterogeneity: the turnover acts as a device to filter permanent, but difficult-to-observe heterogeneity.

[10] The definition of g is slightly more complex now as it involves an expectation over types of all other workers, but we assume it has the same properties.

Evidently, ψ^* depends on the nature of the production function, as represented by $g(\cdot)$. When the key to profitable production is that performance of the worker does not fall significantly below a particular threshold (i.e., the work is *critical* to production), $g(\cdot)$ will be a function that drops off very rapidly when performance is below the desired threshold. We provide a more precise definition of "more critical," and then consider how this aspect of technology shapes the optimal low-powered incentive contract.

Definition: *Let a be a parameter that alters the shape of g (while maintaining its differentiability and concavity). Let p_1^* and p_2^* be the equilibrium levels of performance for type 1 and type 2 workers when $a = a_0$. Then increasing a from a_0 makes performance* **more critical** *if*

$$g_a(p) \text{ is } \begin{cases} = 0, & \text{for } p = p_1^* \text{ or } p = p_2^* \\ > 0, & \text{for } p \in (p_1^*, p_2^*) \\ < 0, & \text{for } p \notin [p_1^*, p_2^*] \end{cases}$$

and

$$\theta g_{ap}(p_1^*) > -(1-\theta) g_{ap}(p_2^*) > 0. \tag{12}$$

The idea behind this definition is to generate changes in the production function like that shown in Figure 1, with marginal product rising at p_1^* and falling at p_2^*, making it more and more critical that the worker supply performance close to 1.[11,12] In Figure 1, a firm with technology $\widetilde{g}$ is more averse to downward variation in performance, while upward variation confers less advantage.

<< Insert Figure 1 about here >>

[11] We hold scale fixed in order to isolate the effects of changing the shape of the production function, although the exact way in which we do so is somewhat arbitrary. The same results can be obtained using a definition that bends the production function down around $g(1)$ rather than forcing it through p_1^* and p_2^*.

[12] The critical worker concept is related to, but not identical to, complementarity. As an illustration, suppose that $G(\cdot)$ is a constant-elasticity-of-substitution function. Consider a firm seeking to motivate a worker, given that all other workers provide $\bar{p} = 1$. Then $g(p) = G(p, \bar{p}, \ldots, \bar{p}) = A[p^{-a} + (n-1)]^{-\frac{1}{a}}$, with A chosen so that $g'(1) = 1$. The shape of $g(p)$ can be changed in the way specified above by moving the elasticity parameter a toward ∞, while changing A to keep the scale from changing.

Condition (12) is less restrictive than it first appears. If g changes only slightly near p_1^* and θ is tiny, the problem of motivating type 1 workers takes a back seat to motivating type 2 workers. Thus for small changes in the technology something like condition (12) is needed for our subsequent results. But large enough changes in the technology automatically satisfy the condition, essentially because $g'(p)$ is bounded below by 0 for $p > p_2^*$, while there is no such bound for $p < p_1^*$. Our key result concerning critical workers and low-powered incentives is Proposition 5.

Proposition 5: *Suppose that f obeys the monotone likelihood ratio property. Then making performance more critical (increasing a) has the following consequences.*

- *i.* p_1^* *increases* ($dp_1^*/da > 0$).
- *ii.* p_2^* *either falls or rises by less than* p_1^* ($dp_2^*/da < dp_1^*/da$).
- *iii. If* $\theta g_{ap}(p_1^*)$ *is sufficiently large relative to* $|(1-\theta)g_{ap}(p_2^*)|$, *then* w^* *rises* ($dw^*/da > 0$).
- *iv. If* $\theta g_{ap}(p_1^*)$ *is sufficiently close to* $|(1-\theta)g_{ap}(p_2^*)|$, *then* $\bar{x}^*$ *falls* ($d\bar{x}^*/da > 0$).
- *v. The probability of termination falls for both types.*

A proof of Proposition 5 can be found in the appendix. Under low-powered incentives, performance is motivated solely by the possibility of losing a valuable job. Intuition might suggest that when performance is critical—when a production-line worker can ruin a high-value product with poor performance, for instance—the firm is likely to experience higher turnover as a byproduct of its attempt to ensure adequate performance. In fact, the present model of low-powered incentives predicts the opposite outcome: When performance is critical, firms adopt policies that result in a high retention rate.

Proposition 4 provides the appropriate intuition for this result. When response to incentives is heterogenous, the optimal dismissal threshold $\bar{x}$ results in termination probabilities $F(\bar{x}^* - p_j^*)$ that are "too low" to optimize the performance of a high-productivity worker, and "too high" to optimize performance of a low-productivity worker. If the job were more critical, the firm would prefer an employment policy tilted more toward optimizing performance of the low-productivity worker. (Note that the key quantity here is the termination probability, not the dismissal threshold; the firm's goal in manipulating $\bar{x}$ is to change the termination probabilities.)

To accomplish an increase in $\hat{p}_1$, the firm must lower $\bar{x}$ and/or raise w. The firm would like to accomplish its adjustment solely through $\bar{x}$, as this is costless. That would raise $\hat{p}_1$ and lower $\hat{p}_2$. Holding all else fixed, $\hat{p}_1$ increases for low values of $\bar{x}$ (as the threat of termination becomes large enough to matter), but then reaches a maximum and declines. Proposition 4 says that p_1^* is to the right of its maximum, and this maximum bounds the firm's willingness to increase p_1^* by adjusting on only the $\bar{x}$ margin: As $\partial \hat{p}_1 / \partial \bar{x}$ nears zero, $\partial \hat{p}_2 / \partial \bar{x}$ does not. Therefore, w^* must increase if the movement toward critical performance is rapid enough, i.e., $g_{ap}(p_1)$ is large relative to $|g_{ap}(p_2)|$.

IV. Internal Competition and Low-Powered Incentives

It is tempting to view low-powered incentives as a fall-back strategy used when high-powered incentives are not available or are restricted in some way. In fact, in this section we show that the interaction between the two is more complex than this portrayal; in some circumstances low-powered incentives can be an important mechanism for dealing with problems that arise from the use of higher-powered incentives. This section studies these issues by supplementing low-powered incentives with a rank-order tournament similar to that of Lazear and Rosen (1981).

As mentioned earlier, a compensation scheme based directly on x (which is private and subjective) would be easily manipulated by the employer. However, Malcomson (1984) and others have observed that if the size of the prize pool is contractible, rank-order tournaments with preannounced prizes can circumvent this form of moral hazard.

The tournament is structured as follows. Two employees, who are observed by the same manager, compete to win a prize z. Each period's compensation consists of the prize and a base wage w. The prize is awarded to the employee observed to have the highest value of x. The manager is honest in awarding the prize.

The functioning of rank-order tournaments in the case of homogenous workers has been extensively studied.[13] One way to tell the story is this: Suppose the firm offers employees fixed expected compensation, at a level high enough that workers are willing to take the job for some anticipated positive level of performance (i.e., $\mathrm{E}\{V\} \geq V^a$ when performance is 0). If the firm sets $z = 0$ there is no incentive, and employees supply $p = 0$. As a thought experiment, let the firm increase z, keeping overall compensation fixed. The competition for z provides an incentive to increase performance. Thus workers move closer to their participation constraints. If the firm has chosen the level of expected compensation optimally, when the firm has increased the prize to a point that induces the efficient level of performance, workers' participation constraints also bind.

A. The model augmented by a rank-order tournament

It is not difficult to generalize the low-powered incentives model to incorporate a rank-order tournament between two workers. So that the model has an equilibrium in the special case (only) where the firm uses no low-powered incentives, it is necessary to complicate the model by assuming that disutility is a convex function of performance, $\eta_j c(p)$. We continue to assume $\eta_1 > \eta_2$, $\mathrm{E}\{\eta\} = 1$, and $\Pr\{\eta_1\} = \theta$. The employment policy is $\psi = \{w, z, \bar{x}\}$, where w is the base wage.

Let $P(p; p'_1, p'_2)$ be the probability of winning the prize by supplying performance p conditional on one's opponent supplying either p'_1 or p'_2, depending on her

[13] Textbook treatments include Lazear (1998) and Gibbons (1992).

type. Then

$$P(p;p'_1,p'_2) = 1-\theta\int_{-\infty}^{\infty} F(\epsilon+p'_1-p)f(\epsilon)\,d\epsilon - (1-\theta)\int_{-\infty}^{\infty} F(\epsilon+p'_2-p)f(\epsilon)\,d\epsilon.$$

We will restrict attention to symmetric Bayesian Nash equilibria, with equilibrium performance levels denoted $\{\hat{p}_1, \hat{p}_2\}$.

Suppose that both the employee and her opponent supply $\hat{p}_1$ and $\hat{p}_2$ in the future and that the opponent supplies $\hat{p}_1$ or $\hat{p}_2$ now. If the employee's current type is j, she chooses p_j to maximize

$$V(p_j) = w+P(p_j;\hat{p}_1,\hat{p}_2)z-\eta_j c(p_j)+F(\bar{x}-p_j)\beta V^a+[1-F(\bar{x}-p_j)]\beta \mathrm{E}_\eta\{V\}. \quad (13)$$

The first-order condition for (13) is

$$0 = P_1(p_j;\hat{p}_1,\hat{p}_2)z-\eta_j c'(p_j)+f(\bar{x}-p_j)\beta[\mathrm{E}_\eta\{V\}-V^a]. \quad (14)$$

As noted above, in standard tournament models (with a homogenous workforce), increasing z always increases equilibrium performance (as long as the participation constraint does not bind). The key to understanding the relationship between low- and high-powered incentives is to ask whether the same is true here: Does an increase in z always increase performance when worker productivity is heterogeneous?

We analyze the same thought experiment described above: changing z, while holding overall compensation constant. Let overall compensation for both employees be $2W$. Since the tournament is fair, the base wage is $w(z) = W - z/2$. Differentiating equation (14) with respect to z, then imposing the equilibrium conditions, $dp_j/dz = d\hat{p}_j/dz$ for $j = 1, 2$, yields

$$\frac{d\hat{p}_1}{dz} = \frac{P_1(\hat{p}_1;\hat{p}_1,\hat{p}_2) - z(1-\theta)A\frac{d\hat{p}_2}{dz} + f(\bar{x}-\hat{p}_1)\beta\frac{d\mathrm{E}_\eta\{V\}}{dz}}{-z(1-\theta)A+\eta_1 c''(\hat{p}_1)+f'(\bar{x}-\hat{p}_1)\beta[\mathrm{E}_\eta\{V\}-V^a]}, \quad (15)$$

$$\frac{d\hat{p}_2}{dz} = \frac{P_1(\hat{p}_2;\hat{p}_1,\hat{p}_2) + z\theta A\frac{d\hat{p}_1}{dz} + f(\bar{x}-\hat{p}_2)\beta\frac{d\mathrm{E}_\eta\{V\}}{dz}}{z\theta A+\eta_2 c''(\hat{p}_2)+f'(\bar{x}-p_2)\beta[\mathrm{E}_\eta\{V\}-V^a]}, \quad (16)$$

where $A = \int_{-\infty}^{\infty} f'(\epsilon + \hat{p}_1 - \hat{p}_2) f(\epsilon)\, d\epsilon > 0$.[14]

Second-order conditions imply that the denominators are positive. The equilibrium impact of an increase in the power of the incentive (an increase in z), for performance in both the low-productivity state ($\hat{p}_1$) and high-productivity state ($\hat{p}_2$), is determined by the three terms in the numerator.

The first term corresponds to the "direct effect" on performance of increasing the prize—the effect of changing one's own performance on the probability of winning. When a worker provides higher performance, the prospects of winning the prize improve, so this term is positive.

The second term can be thought of as an "indirect effect" created by the changes in the performance of competitors whose productivity state is different from one's own. (The term corresponding to competitors of one's own type is always zero.) This term is central to our subsequent analysis, especially of performance in the low-productivity state $\hat{p}_1$. The term has opposite signs in (15) and (16); increased performance by a type 1 competitor motivates more effort from a type 2 worker, but increased performance by a type 2 competitor *discourages* a type 1 employee. This effect, which we term the *uneven playing field effect,* is based on appealing intuition: When a more able competitor decides to work harder, the marginal expected value of one's own performance deteriorates.

The third term reflects the changed value of the job induced by changes in future performance. By construction, expected compensation has not changed, but because performance levels change, there is an indirect effect on the value of the job. Any change that reduces the value of the job in turn reduces the impact of incentives from competition if the threat of dismissal is part of the firm's personnel practice.

---

[14] Derivation of these equations is quite tedious. Details are available from the authors.

B. The tournament alone

With these observations in mind, we first explore how the tournament functions in the absence of low-powered incentives. That is, we ask how the tournament operates when the firm relies only on internal competition, not on the threat of dismissal. Proposition 6 follows from setting $\bar{x} = -\infty$ in equations (15) and (16) (making the model comparable to Lazear and Rosen's, but with heterogeneity):

Proposition 6: *Suppose the firm uses exclusively a rank-order tournament and the participation constraint $\mathrm{E}_\eta\{V\} \geq V^a$ does not bind. Then changes to the size of the prize z, holding overall compensation fixed, have the following consequences.*

(i) *High-productivity workers always respond positively to increasing the size of the tournament, that is, $d\hat{p}_2/dz > 0$ for all values of z.*

(ii) *For a sufficiently small prize, low-productivity workers also respond positively: $d\hat{p}_1/dz > 0$. Starting from larger prizes, increasing z can de-motivate low-productivity workers, that is, $d\hat{p}_1/dz \leq 0$ is possible.*

(iii) *If the prize is sufficiently small and $\theta \leq 1/2$, type 1 workers respond less vigorously to the tournament: $d\hat{p}_1/dz < d\hat{p}_2/dz$.*

Proof: Setting $\bar{x} = -\infty$ eliminates the terms involving f and f'. Part (i): The signs of the remaining terms in the numerators imply that $d\hat{p}_2/dz > 0$. Part (ii): If z is sufficiently small, the first term in the numerator of $d\hat{p}_1/dz$ dominates. The possibility of $d\hat{p}_1/dz \leq 0$ we demonstrate by example below. Part (iii): $P_1(\hat{p}_1; \hat{p}_1, \hat{p}_2) \leq P_1(\hat{p}_2; \hat{p}_1, \hat{p}_2)$ if and only if $\theta \leq 1/2$. If z is sufficiently small, $\hat{p}_1$ and $\hat{p}_2$ are near enough to zero to ensure that $\eta_1 c''(\hat{p}_1) \geq \eta_2 c''(\hat{p}_2)$. ∎

Figure 2 illustrates the claims of Proposition 6 using a simple simulation of the thought experiment discussed earlier—increasing the prize z while holding overall compensation fixed. As indicated by the proposition, $\hat{p}_2$ increases with the prize up to the point where the participation constraint binds (at $z = 9.94$). For relatively small tournaments, $\hat{p}_1$ is also increasing in z, but it rises more slowly than does $\hat{p}_2$. More importantly, $\hat{p}_1$ declines as the prize grows large, a consequence of the uneven playing field effect. Therefore, *if the compensation scheme is being designed* *to motivate critical workers, it may not be optimal to push the workers to their participation constraint.*

< Insert Figure 2 about here. >

A final observation is in order. In this example, the base wage is negative for $z > 9$. This will restrict the scope of the tournament if the firm cannot implement contracts that call for upfront transfers from workers to the firm.

C. Tournaments with low-powered incentives

Figure 3 illustrates how low-powered incentives change the picture. Figure 3 combines the simulation shown in Figure 2 with a parallel solution that uses the hybrid model (low-powered incentives and a tournament) for an arbitrary value of $\bar{x}$.[15] The two sets of solutions necessarily converge as they approach the participation constraint. The mechanism that creates this convergence is intuitive: the low-powered incentives model differs from the pure tournament model in that the asset value of the job is put at risk. Just as in the pure tournament model, that value declines toward 0 as z increases.

< Insert Figure 3 about here. >

Three features of Figure 3 stand out. First, this example dramatically illustrates the uneven playing field effect and its importance when the job is critical. The performance of a low-productivity worker is higher under an employment policy that uses only low-powered incentives ($z = 0$) than for a policy that uses only a tournament of *any* size.

A second, not surprising, but nevertheless important, feature of Figure 3 is that the use of terminations increases both $\hat{p}_1$ and $\hat{p}_2$ for every value of z. For most values of z, workers of both types respond less vigorously to changes in z when low-powered

[15] The optimal value of $\bar{x}$ depends on g and z, but our suboptimal choice of $\bar{x}$ adequately illustrates the features of low-powered incentives.

incentives are in place. This is simply a consequence of the fact that the asset value of the job—which low-powered incentives exploit—declines as z increases, offsetting the direct effect of competition. On the downward-sloping part of the low-powered incentives curve for $\hat{p}_1$, this effect reinforces the uneven playing field effect, causing a steeper drop-off in performance than when low-powered incentives are not used.

Third, the firm using exclusively low-powered incentives ($z = 0$) benefits by introducing a tournament (this is easy to see from equations (15) and (16) as well). How far should the low-powered incentives firm push the tournament? For some parameter values, both of the "tournament + LPI" lines in Figure 3 would be strictly increasing until the participation constraint is reached. In that case, assuming the tournament faced none of the impediments mentioned in the literature, the optimal employment policy with low-powered incentives would be equivalent to the pure tournament model—raise z until the participation constraint is satisfied (at which point $\bar{x}$ is irrelevant). With the parameters used in Figure 3, however, the low-powered incentives curve has a downward-sloping segment. In such a case, the answer depends on the shape of $g(p)$. If performance is sufficiently critical, it is optimal to stop short of the participation constraint. In other words, if performance is sufficiently critical, for some z at which the participation constraint does not bind, an increase in the high-powered component of compensation serves only to reduce expected contribution to firm value, that is,[16]

$$\theta g'(\hat{p}_1)\frac{d\hat{p}_1}{dz} + (1-\theta)g'(\hat{p}_2)\frac{d\hat{p}_2}{dz} < 0.$$

Further increases in z elicit higher p_2 at the expense of lower p_1.

V. Conclusion

The arrangements employers typically reach with their workforce look quite different than incentive contracts derived by economic theorists. Many employees are

[16] Recall that $d\hat{p}_j/dz$ does not depend on g.

hired with an understanding that they will be offered a fixed level of compensation, along with continued employment, so long as their performance appears to exceed some minimal threshold. Thus Prendergast (1999) argues, "A critical avenue for future research should be to better understand the evaluation and compensation of those with non-contracted output."

In that spirit, this paper provides a parsimonious characterization of compensation, performance, and turnover policy for firms that use low-powered incentives to solve agency problems. Our analysis centers on workplaces that share two features: First, managers rely on private subjective information in evaluating worker performance. Second, there is unobservable productivity variation among workers. For such settings we contrast the properties of low-powered incentives with a potentially viable high-powered alternative—a rank-order tournament with pre-announced, verifiable prizes.

We find that the effectiveness of any incentive scheme hinges crucially on the workplace technology, in particular on the extent to which the firm can tolerate downward variation in performance. When some workers are *critical*—a concept closely related to Kremer's o-ring technology—the value of the firm's output varies significantly with the poorest performance in the workforce. Low-powered incentives are effective in this environment, because they can motivate workers who would otherwise be most inclined to provide the lowest performance. In this model, a firm that uses low-powered incentives to motivate workers responds to a change in technology that makes workers "more critical" by shifting toward policies with relatively low worker turnover and high compensation.

With low-powered incentives, workers perform well so as to avoid losing valuable jobs. Making jobs valuable is costly to firms. Although firms cannot base pay directly on performance (because managers' assessments of performance are subjective and private), they may be able to implement incentives based on internal competition and thereby reduce costs. We show that in fact a firm that uses low-powered incentives will reduce employment costs if it can introduce a rank-order tournament with a small prize. We also show, however, that firms employing critical workers will not generally abandon low-powered incentives in favor of alternative tournament-based incentives. Internal competition interacts with heterogeneity in productivity by inducing a de-motivating "uneven playing field" effect—since it is harder for low-productivity workers to win the prize, they do not try as hard. The uneven playing field effect partly or fully offsets the direct effect of competition, and its importance is magnified if the worker's performance is critical. In this situation, a large prize can be counterproductive.

In many work environments, low-powered incentives appear to have a further advantage, which we do not model here. If the heterogeneity in productivity is persistent, the use of terminations for particularly poor observed performance tends to weed out low-productivity workers, so the terminations do double duty as an incentive device and a means of screening. Again, this feature is especially important when performance is critical. Fully understanding this aspect of low-powered incentives would require careful treatment of the process by which the firm learns about workers' types.

Appendix

Proofs of Propositions 4 and 5

Proof of Proposition 4:

Step 1. Differentiation and simplification. Equations (10) implicitly define best responses $\hat{p}_1$ and $\hat{p}_2$. Note that $Q(\bar{x}, w)$ does not depend on j, and

$$
\begin{aligned}
Q_w(w,\bar{x}) &= -\theta f(\bar{x}-\hat{p}_1)\frac{\partial \hat{p}_1}{\partial w} - (1-\theta) f(\bar{x}-\hat{p}_2)\frac{\partial \hat{p}_2}{\partial w} \\
Q_{\bar{x}}(w,\bar{x}) &= \theta f(\bar{x}-\hat{p}_1)\left(1-\frac{\partial \hat{p}_1}{\partial \bar{x}}\right) + (1-\theta) f(\bar{x}-\hat{p}_2)\left(1-\frac{\partial \hat{p}_2}{\partial \bar{x}}\right).
\end{aligned}
$$

Now for $j=1$, differentiate (10) with respect to w:

$$
\begin{aligned}
1 = \theta\eta_1\frac{\partial \hat{p}_1}{\partial w} &+ (1-\theta)\eta_2\frac{\partial \hat{p}_2}{\partial w} + Q_w(\bar{x},w)\left(\frac{\eta_1}{f(\bar{x}-\hat{p}_1)}\right) \\
&+ Q(\bar{x},w)\frac{\eta_1 f'(\bar{x}-\hat{p}_1)}{f(\bar{x}-\hat{p}_1)^2}\frac{\partial \hat{p}_1}{\partial w}.
\end{aligned} \tag{A1}
$$

And with respect to $\bar{x}$,

$$
\begin{aligned}
0 = \theta\eta_1\frac{\partial \hat{p}_1}{\partial \bar{x}} &+ (1-\theta)\eta_2\frac{\partial \hat{p}_2}{\partial \bar{x}} + Q_{\bar{x}}(\bar{x},w)\left(\frac{\eta_1}{f(\bar{x}-\hat{p}_1)}\right) \\
&- Q(\bar{x},w)\frac{\eta_1 f'(\bar{x}-\hat{p}_1)}{f(\bar{x}-\hat{p}_1)^2}\left(1-\frac{\partial \hat{p}_1}{\partial \bar{x}}\right).
\end{aligned} \tag{A2}
$$

Substitute for Q_w in (A1) and apply equation (9):

$$
\begin{aligned}
1 &= \theta\eta_1\frac{\partial \hat{p}_1}{\partial w} + (1-\theta)\eta_2\frac{\partial \hat{p}_2}{\partial w} - \frac{\eta_1}{f(\bar{x}-\hat{p}_1)}\left[\theta f(\bar{x}-\hat{p}_1)\frac{\partial \hat{p}_1}{\partial w} + (1-\theta) f(\bar{x}-\hat{p}_2)\frac{\partial \hat{p}_2}{\partial w}\right] \\
&\quad + Q(\bar{x},w)\frac{\eta_1 f'(\bar{x}-\hat{p}_1)}{f(\bar{x}-\hat{p}_1)^2}\frac{\partial \hat{p}_1}{\partial w} \\
&= \theta\eta_1\frac{\partial \hat{p}_1}{\partial w} + (1-\theta)\eta_2\frac{\partial \hat{p}_2}{\partial w} - \eta_1\theta\frac{\partial \hat{p}_1}{\partial w} - (1-\theta)\eta_1\frac{\eta_2}{\eta_1}\frac{\partial \hat{p}_2}{\partial w} \\
&\quad + Q(\bar{x},w)\frac{\eta_1 f'(\bar{x}-\hat{p}_1)}{f(\bar{x}-\hat{p}_1)^2}\frac{\partial \hat{p}_1}{\partial w} \\
&= Q(\bar{x},w)\frac{\eta_1 f'(\bar{x}-\hat{p}_1)}{f(\bar{x}-\hat{p}_1)^2}\frac{\partial \hat{p}_1}{\partial w}.
\end{aligned} \tag{A3}
$$

Equation (A3) implies that $\partial \hat{p}_j / \partial w > 0$. Applying essentially the same algebra to equation (A2) yields:

$$1 = Q(\bar{x}, w) \frac{\eta_1 f'(\bar{x} - \hat{p}_1)}{f(\bar{x} - \hat{p}_1)^2} \left(1 - \frac{\partial \hat{p}_1}{\partial \bar{x}} \right). \tag{A4}$$

Comparing (A3) and (A4) we have

$$\frac{\partial \hat{p}_1}{\partial w} = 1 - \frac{\partial \hat{p}_1}{\partial \bar{x}}. \tag{A5}$$

An analogue for $j = 2$ is derived in the same way.

Step 2. Application of MLRP. By strict MLRP,

$$0 < \frac{\partial}{\partial \overline{x}} \left[\frac{f(\overline{x} - \hat{p}_2)}{f(\overline{x} - \hat{p}_1)} \right] = \frac{f(\overline{x} - \hat{p}_1) f'(\overline{x} - \hat{p}_2) - f(\overline{x} - \hat{p}_2) f'(\overline{x} - \hat{p}_1)}{f(\overline{x} - \hat{p}_1)^2}$$

so, therefore,

$$\frac{f(\overline{x} - \hat{p}_1)}{f(\overline{x} - \hat{p}_2)} > \frac{f'(\overline{x} - \hat{p}_1)}{f'(\overline{x} - \hat{p}_2)}.$$

Differentiating (9) w.r.t. w gives

$$\eta_2 f'(\overline{x} - \hat{p}_1) \frac{\partial \hat{p}_1}{\partial w} = \eta_1 f'(\overline{x} - \hat{p}_2) \frac{\partial \hat{p}_2}{\partial w}.$$

Rearranging and applying the MLRP,

$$\frac{\partial \hat{p}_2 / \partial w}{\partial \hat{p}_1 / \partial w} = \frac{f'(\overline{x} - \hat{p}_1)}{f'(\overline{x} - \hat{p}_2)} \frac{\eta_2}{\eta_1} < \frac{f(\overline{x} - \hat{p}_1)}{f(\overline{x} - \hat{p}_2)} \frac{\eta_2}{\eta_1} = 1.$$

Thus $\partial \hat{p}_1 / \partial w > \partial \hat{p}_2 / \partial w$.

Step 3. Implication of profit maximization. The first-order condition for $\overline{x}$ is

$$0 = \theta g'\big(\hat{p}_1(\overline{x}^*, w^*)\big) \frac{\partial \hat{p}_1^*}{\partial \overline{x}} + (1 - \theta) g'\big(\hat{p}_2(\overline{x}^*, w^*)\big) \frac{\partial \hat{p}_2^*}{\partial \overline{x}},$$

which implies that $\partial \hat{p}_1 / \partial \overline{x}$ and $\partial \hat{p}_2 / \partial \overline{x}$ have opposite signs or are both zero at ψ^*. However, $\partial \hat{p}_1^* / \partial \overline{x} = \partial \hat{p}_2^* / \partial \overline{x} = 0$ together with equation (A5) implies that $\partial \hat{p}_1^* / \partial w = \partial \hat{p}_2^* / \partial w = 1$, which the MLRP rules out. Therefore, $\partial \hat{p}_1^* / \partial \overline{x}$ and $\partial \hat{p}_2^* / \partial \overline{x}$ have opposite signs, and equation (A5) implies $\partial p_1^* / \partial \overline{x} < 0$ and $\partial \hat{p}_1^* / \partial w > 1$. ∎

Proof of Proposition 5: Part (i): Prove that $dp_1^*/da > 0$. Note that changing a does not change the feasible combinations of p_1, p_2, w, and $\bar{x}$. In particular, as a increases, profits are unchanged at ψ^*, which still induces (p_1^*, p_2^*). Suppose that $dp_1^*/da < 0$. Then $dp_2^*/da > 0$ is necessary to satisfy equation (11). Since the hypothesized change was feasible, it must not have been profitable before; the change in profits in the direction of $dp_1^*/da < 0$ and $dp_2^*/da > 0$ must not be positive:

$$\theta g_p(p_1^*)\frac{dp_1^*}{da} + (1-\theta)g_p(p_2^*)\frac{dp_2^*}{da} - \frac{dw^*}{da} \leq 0.$$

But differentiating the left-hand side this inequality, holding the direction of change fixed, shows that this direction is even less profitable as a increases:

$$\theta g_{ap}(p_1^*)\frac{dp_1^*}{da} + (1-\theta)g_{ap}(p_2^*)\frac{dp_2^*}{da} < 0.$$

Therefore it must be the case that $dp_1^*/da > 0$.

Part (iii): Let $\Pi^* = \theta g(p_1^*) + (1-\theta)g(p_2^*) - w^*$. The first-order conditions for profit maximization are $0 = \Pi_w^*$ and $0 = \Pi_{\bar{x}}^*$. Differentiating them with respect to a produces the following comparative statics results:

$$\frac{dw^*}{da} = \frac{\Pi_{a\bar{x}}^*\Pi_{w\bar{x}}^* - \Pi_{aw}^*\Pi_{\bar{x}\bar{x}}^*}{\Pi_{ww}^*\Pi_{\bar{x}\bar{x}}^* - (\Pi_{w\bar{x}}^*)^2},$$

$$\frac{d\bar{x}^*}{da} = \frac{\Pi_{aw}^*\Pi_{w\bar{x}}^* - \Pi_{a\bar{x}}^*\Pi_{ww}^*}{\Pi_{ww}^*\Pi_{\bar{x}\bar{x}}^* - (\Pi_{w\bar{x}}^*)^2},$$

Second-order conditions guarantee that the denominators are positive and that Π_{ww} and $\Pi_{\bar{x}\bar{x}}$ are negative. Therefore the numerator of dw^*/da is positive if Π_{aw} is large enough relative to $|\Pi_{a\bar{x}}|$. However,

$$\Pi_{aw}^* + \Pi_{a\bar{x}}^* = \Pi_{aw}^* - |\Pi_{a\bar{x}}^*| = \theta g_{ap}(p_1^*) + (1-\theta)g_{ap}(p_2^*).$$

so the condition required by the proposition is equivalent to Π_{aw}^* large relative to $|\Pi_{a\bar{x}}^*|$. Note that the quantity on the right-hand side can be arbitrarily large without affecting any terms in dw^*/da other than Π_{aw}^* and $\Pi_{a\bar{x}}^*$.

Part (iv): Under the stated condition, $\Pi^*_{aw} \approx |\Pi^*_{a\bar{x}}|$, so the numerator of $d\bar{x}^*/da$ is approximately $\Pi^*_{aw}(\Pi^*_{w\bar{x}} + \Pi^*_{ww})$. Some algebra shows that

$$\Pi^*_{w\bar{x}} + \Pi^*_{ww} = \theta g_{pp}(p_1^*)\frac{\partial \hat{p}_1}{\partial w} + (1-\theta)g_{pp}(p_2^*)\frac{\partial \hat{p}_2}{\partial w} < 0.$$

Therefore, $d\bar{x}^*/da < 0$ under the stated condition.

Part (v): The probability of termination for type j is $F(\bar{x}^* - \hat{p}_j^*)$, and

$$\begin{aligned}\frac{d}{da}F(\bar{x}^* - p_j^*) &= f(\bar{x}^* - \hat{p}_j^*)\left[\frac{d\bar{x}^*}{da} - \frac{\partial p_j^*}{\partial \bar{x}}\frac{d\bar{x}^*}{da} - \frac{\partial p_j^*}{\partial w}\frac{dw^*}{da}\right]\\ &= f(\bar{x}^* - \hat{p}_j^*)\frac{\partial p_j^*}{\partial w}\left[\frac{d\bar{x}^*}{da} - \frac{dw^*}{da}\right].\end{aligned}$$

The second equality follows from Proposition 4. Thus $dF(\bar{x}^* - p_j^*)/da$ has the same sign as $dx^*/da - dw^*/da$. Now

$$\begin{aligned}\frac{d\bar{x}^*}{da} - \frac{dw^*}{da} &= \frac{-\Pi^*_{a\bar{x}}[\Pi^*_{ww} + \Pi^*_{w\bar{x}}] + \Pi^*_{aw}[\Pi^*_{\bar{x}\bar{x}} + \Pi^*_{w\bar{x}}]}{\Pi^*_{ww}\Pi^*_{\bar{x}\bar{x}} - (\Pi^*_{w\bar{x}})^2}\\ &= \frac{\theta(1-\theta)\big(g_{ap}(p_1^*)g_{pp}(p_2^*) - g_{ap}(p_2^*)g_{pp}(p_1^*)\big)\left(\frac{\partial \hat{p}_1^*}{\partial w} - \frac{\partial \hat{p}_1^*}{\partial w}\right)}{\Pi^*_{ww}\Pi^*_{\bar{x}\bar{x}} - (\Pi^*_{w\bar{x}})^2}. \quad \text{(A6)}\end{aligned}$$

The second equality uses the fact that

$$\frac{\partial^2 \hat{p}_j}{\partial w^2} = \frac{\partial^2 \hat{p}_j}{\partial w^2} = \frac{\partial^2 \hat{p}_j}{\partial w \partial \bar{x}},$$

which follows from Proposition 4. The definition of "more critical" implies that $g_{ap}(p_1^*) > 0$ and $g_{ap}(p_2^*) < 0$. Finally, $\partial \hat{p}_1^*/\partial w > \partial \hat{p}_2^*/\partial w$ (Proposition 4). Thus the numerator is negative, as required.

Part (ii): From the proof of part (v) and Proposition 4,

$$\frac{d\bar{x}^*}{da} - \frac{dp_1^*}{da} = \frac{\partial p_1^*}{\partial w}\left[\frac{d\bar{x}^*}{da} - \frac{dw^*}{da}\right] < \frac{\partial p_2^*}{\partial w}\left[\frac{d\bar{x}^*}{da} - \frac{dw^*}{da}\right] = \frac{d\bar{x}^*}{da} - \frac{dp_2^*}{da}$$

It follows that $dp_1^*/da > dp_2^*/da$. ∎

References

Baker, George, Robert Gibbons, and Kevin J. Murphy. "Subjective Performance Measures and Optimal Incentive Contracts," *Quarterly Journal of Economics,* November 1994, 109(4), pp. 1125-56.

Gibbons, Robert. *Game Theory for Applied Economists,* 1992, Princeton: Princeton University Press.

Gibbons, Robert. "Incentives in Organizations," *Journal of Economic Perspectives,* Fall 1998, 12(4), pp. 115-32

Kremer, Michael. "The O-Ring Theory of Economic Development," *Quarterly Journal of Economics,* August 1993, 108(3), pp. 551-75.

Lazear, Edward P. "Why is There Mandatory Retirement?" *Journal of Political Economy,* December 1979, 87(6), pp. 721-35.

Lazear, Edward P. "Pay Equality and Industrial Politics," *Journal of Political Economy,* June 1989, 97(3), pp. 561-80.

Lazear, Edward P. *Personnel Economics for Managers,* 1998, New York: John Wiley and Sons.

Lazear, Edward P. and Sherwin Rosen. "Rank-Order Tournaments as Optimum Labor Contracts," *Journal of Political Economy,* October 1981, 89(5), pp. 841-64.

MacLeod, W. Bentley and Daniel Parent. "Job Characteristics and the Form of Compensation," *Research in Labor Economics,* 1999, 18, pp. 177-242.

Malcomson, James M. "Work Incentives, Hierarchy, and Internal Labor Markets," *Journal of Political Economy,* June 1984, 92(3), pp. 486-507.

Mehta, Shailendra Raj. "The Law of One Price and a Theory of the Firm: A Ricardian Perspective on Interindustry Wages," *Rand Journal of Economics,* Spring 1998, 29(1), pp. 137-56.

Milgrom, Paul R. "Good News and Bad News: Representation Theorems and Applications," *Bell Journal of Economics,* Autumn 1981, 12(2), pp. 380-91.

Prendergast, Canice. "The Provision of Incentives in Firms," *Journal of Economic Literature,* March 1999, 37(1), pp. 7-63.

Prendergast, Canice and Robert H. Topel. "Favoritism in Organizations," *Journal of Political Economy,* October 1996, 104(5), pp. 958-78.

Shapiro, C. and J. Stiglitz. "Involuntary Unemployment as a Worker Discipline Device," *American Economic Review,* June 1984, 74(3) pp. 433-44.

FIGURE 1

Performance is more critical with technology $\tilde{g}(p)$

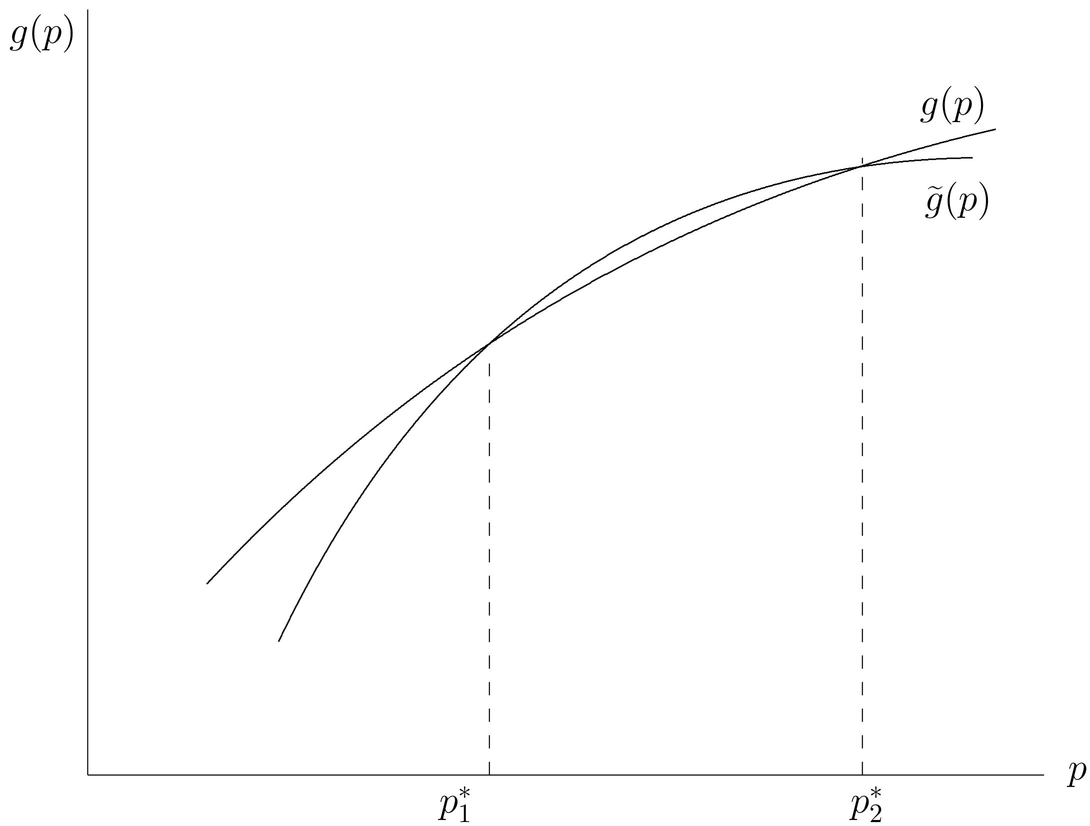
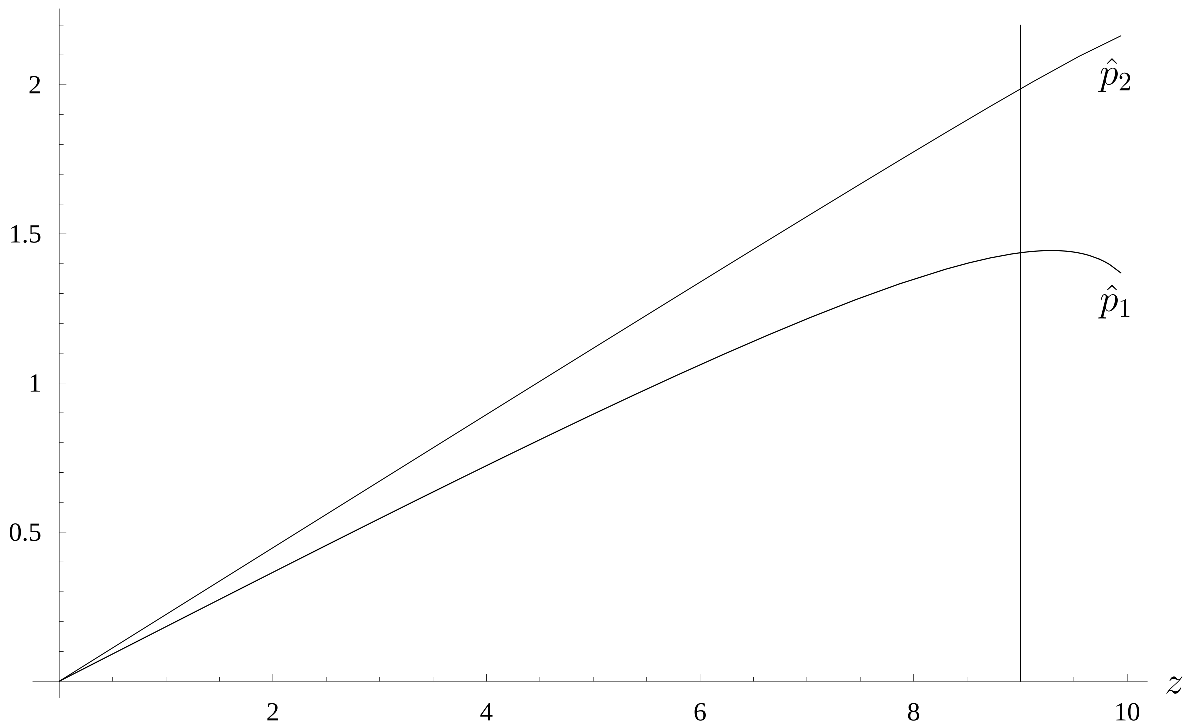


FIGURE 2

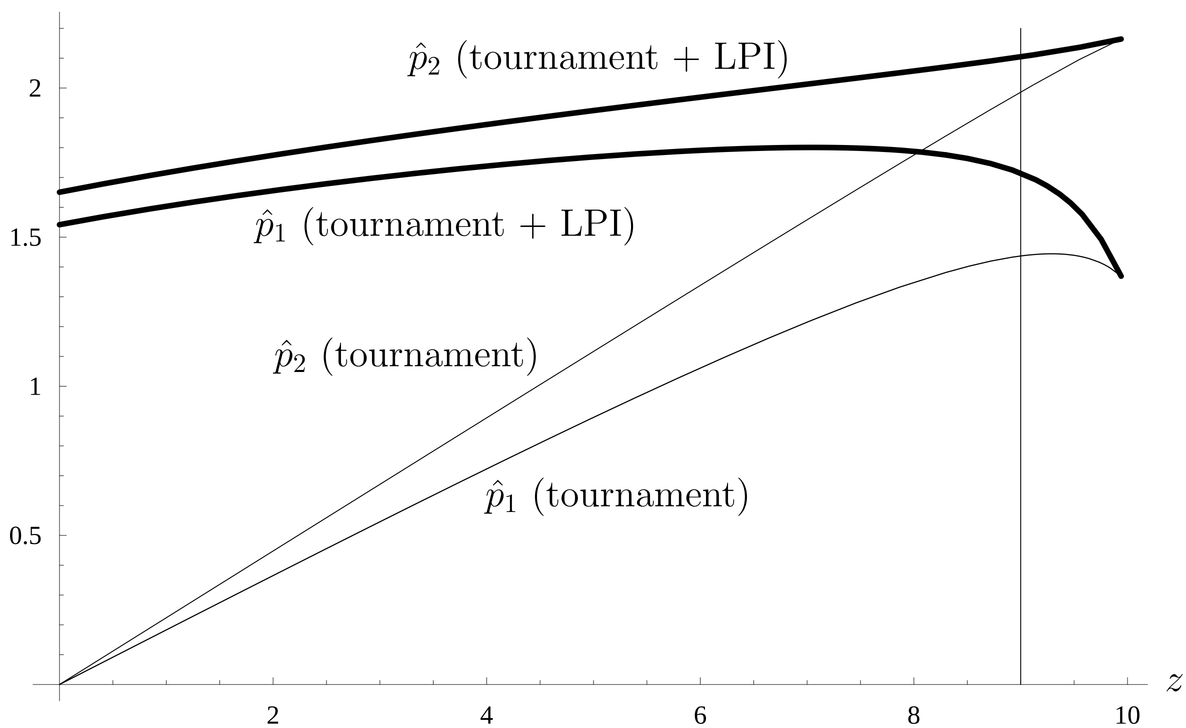
Performance under a rank-order tournament with constant overall compensation



Notes: A separate solution was found for each value of z using equations (14) for $j = 1, 2$ and an analytical expression for $\mathrm{E}_\eta\{V\}$ based on equation (13), setting $\bar{x} = -\infty$ in each equation. The participation constraint is reached at $z = 9.94$. The distribution of ϵ was chosen to be normal with $\sigma = 0.7$, and $c(p) = p^2$. The other parameter values were $\beta = 0.99$, $\theta = 0.1$, $\eta_1 = 1.1$, $\eta_2 = 0.9$, $V^a = 50$. $W = 9/2$ was chosen to generate average performance of 2 at the participation constraint, given the value of V^a.

FIGURE 3

Performance levels in tournament and tournament augmented by low-powered incentives



Notes: The pure tournament curves are replicated from Figure 2. The "tournament + LPI" curves are derived in the same way using $\bar{x} = 1$ and the remainder of the specification unchanged.