Prizes and incentives: towards a general theory of compensation and competition

Barry J. Nalebuff*

and

Joseph E. Stiglitz**

This article analyzes the role of competitive compensation schemes (in which pay depends on relative performance) in economies with imperfect information. These compensation schemes have desirable risk, incentive, and flexibility properties; they provide for an automatic adjustment of rewards and incentives in response to common changes in the environment. When environmental uncertainty is large, such schemes are shown to be preferable to individualistic reward structures; in the limit, as the number of contestants becomes large, expected utility may approach the first-best (perfect information) level. We study the design of contests, including the optimal use of prizes versus punishments and absolute versus relative performance standards. The analysis can also be viewed as a contribution to the multiagent, single-principal problem.

1. Introduction

■ One of the dominant characteristics of modern capitalist economies is the important role played by competition; competition is the force providing work incentives. Rewards within a firm or even to a firm in a competitive marketplace seldom take the form of a pure piece rate (equal to the value of marginal product in the conventional competitive paradigm). More commonly, rewards are at least partially based on *relative* performance: the student with the highest score is given an "*A*," the salesman who sells the most is given a large bonus, the firm with the largest market share becomes the leader, and the best manager gets promoted to the company vice presidency. This conception of competition is better represented by patent races and sports contests than by the static form of pure price competition embodied in the Arrow-Debreu model.

In the past two years, there has been a rekindling of interest in the role of competition as an incentive device. The initial work of Lazear and Rosen (1981) and Stiglitz (1980) has been expanded upon by Green and Stokey (1981), Hart (1981), Holmstrom (1982),

* Harvard University.

** Princeton University.

An earlier version of this article was presented to a conference on the Internal Organization of the Firm, held at the International Institute of Management, Berlin, July 1980. We have benefited greatly from discussions with Sherwin Rosen, Jim Mirrlees, Terence Gorman, Richard Zeckhauser, Kevin Roberts, Ed Lazear, Felix Fitzroy, Partha Dasgupta, Joe Farrell, Jerry Green, Jim Poterba, Roger Koenker, Bengt Holmstrom, Steve Salop, and Hal Varian, and from the comments of participants at the IIM conference and seminars at the Universities of Wisconsin, Michigan, Chicago, Pennsylvania, the Federal Trade Commission, and the London School of Economics, at which earlier versions were presented. We especially appreciate the improvements suggested by Alvin Klevorick and two anonymous referees. Financial support from the National Science Foundation and the Rhodes Trust is gratefully acknowledged.

Mookherjee (1981), Nalebuff (1982), Fitzroy (1981), and O'Keefe, Viscusi, and Zeckhauser (1982). This recent literature has begun to delineate the circumstances in which rewards based on relative output are superior to payments based on individualistic output.

The essential problem with which we are concerned here arises because the input (effort) of workers (managers) is not directly and costlessly observable. Firms must either attempt to monitor inputs or they must devise reward structures in which compensation (using the term in its broadest sense, including promotions, pensions, etc.) is a function of variables which are (a) themselves functions of inputs and (b) less costly to observe than the inputs themselves.

There are three critical characteristics of any reward structure:

(a) *Risk.* Incentive schemes linking rewards to output inevitably result in the workers' bearing risks; and if individuals are risk averse, there is a loss of welfare as a result.

(b) *Incentive levels.* To ameliorate the risk problem, a significant part of most individuals' compensation is not directly related to output. This reduces the incentive to provide the right level of effort and also the incentive to make the correct decisions.

(c) *Flexibility.* The incentive compensation scheme that is "correct" in one situation will not in general be correct in another. In principle, there could be a different incentive structure for each set of environmental variables. Such a contract would obviously be prohibitively expensive to set up; but more to the point, many of the relevant environmental variables are not costlessly observable to all parties to the contract. Thus, a single incentive structure must do in a variety of circumstances. The lack of flexibility of the piece rate system is widely viewed to be its critical shortcoming: the process of adapting the piece rate is costly and contentious.

It is the flexibility characteristic of competitive compensation schemes which makes them so attractive. When a task is easy for one person, it is likely to be easy for his rivals. Consider the effect of a technological innovation that occurs in a work environment in which employees are compensated on the basis of relative performance. Producing more than one's rivals becomes easier so that in the new equilibrium everyone will work harder for the same prizes. Compensation per unit output is automatically reduced. Similarly, while it is very difficult to design a fixed grading system ($90 = A$, $80 = B$, . . .) for a test of unknown difficulty, grading students according to percentiles permits the same scoring system to be used in different courses, different classes, and in different ability groups.

There are a variety of circumstances in which the performance of individuals (or firms) conveys information about the environment which can and should be introduced into the compensation scheme (Hart, 1981; Nalebuff and Stiglitz, 1983; Stiglitz, 1982). Contests are one way of doing this; rewards based on performance standards, which, in turn, are based on the performance of the group, are another.

Compensation structures also serve a second, critical function: they enable the differentiation and screening of workers of different abilities (Stiglitz, 1975). This article focuses on the incentive properties of compensation schemes; hence, except as noted briefly below, we shall assume that all individuals have the same abilities.

We first present the general theory of compensation in Sections 3–5. We then examine the theory of contests (Sections 6–8) and compare contests with other compensation schemes (Sections 9–11). Sections 12–14 provide some important extensions of the basic theory.

2. A summary of the results

■ After setting up the general problem, we focus our attention on certain special cases; for example, in most of the analysis, output is assumed to be linear in effort. We ask under what circumstances would *contests* (where it is only the individual's *rank* that

counts) be preferred to individualistic compensation schemes in which payment is based only on output. Because the analysis is fairly detailed, it may prove useful to summarize the major results here.

(1) The second-best solution will in general require knowledge of each agent's output. This can be simplified when the set of outputs can be characterized through the use of a sufficient statistic. For example, when there is a sufficiently large number of agents, optimal compensation may be based solely on the individual's output and the average output.
(2) With risk neutral agents, the first-best optimum can be obtained through an appropriately designed contest.
(3) In contests with little risk and hence small prizes, no symmetric pure strategy Nash equilibrium will exist.
(4) Although the general principal-agent problem arises because of the difficulty in motivating unmonitored effort, in the second-best solution agents may supply *more* effort than they would in the first-best optimum. Even so, there is a moral hazard problem; working harder would raise utility.
(5) The use of a contest as an incentive device can induce agents to abandon their natural risk aversion and adopt "riskier" and more profitable production techniques.
(6) When there is a large number of contestants: (i) appropriately designed contests can often approach the first-best optimum; and (ii) a penalty to the lowest ranked individual will be superior to a prize to the highest ranked individual in motivating effort.
(7) Contests may be preferred to individualistic reward schemes, especially when the risk associated with common environmental variables is large.
(8) Even when the agents are independent, a two-agent contest can dominate the optimal individualistic linear payment schedule.
(9) It is usually possible to improve on a simple contest by requiring that an agent beat his opponent by a positive gap before being ranked above him in a tournament.
(10) In contests where no pure strategy exists, the mixed strategy equilibrium can come relatively close to the first best when there are only two agents competing. But, as the number of participants grows, competition becomes more intense and eventually all surplus is eliminated.
(11) Tournaments with multiple prizes cannot in general replicate nonlinear compensation schemes (and *vice versa*). In the examples of two commonly used error distributions, even when it is possible to use n different rewards in a contest, the optimal solutions entail at most three different prizes.
(12) When output is no longer linear in effort, a tournament may or may not be able to replicate the marginal incentives corresponding to the first-best allocation of effort. We present an example of each possibility.

3. Basic model

■ We consider a particularly stylized economic environment. Agent i has an observable output, Q_i, which is a random function of his effort, μ_i:

$$Q_i = Q(\mu_i, \theta, \epsilon_i). \tag{1}$$

The randomness arises from a common "environmental" variable, θ, and an "individualistic" noise, ϵ_i. There are several alternative interpretations of (1). In an agricultural context, for instance, θ represents the general weather in the area and ϵ_i the weather (rainfall) on a particular farm. Alternatively, θ could represent the general level of difficulty of a set of tasks, while ϵ_i could represent the individual's comparative advantage (or disadvantage) in performing one of the various tasks. In our analysis it is important that:

(a) $Q_{\mu\theta} \neq 0$, the state of nature affects the return to effort; and

(b) for each θ, from knowledge of Q_i and θ, one cannot infer precisely the value of μ.

Otherwise, (a) the state of nature does not affect the work environment or (b) through observing output it would be possible to directly monitor effort. Throughout, we assume *the contract must be signed before θ and ϵ are known; the worker then observes θ and decides on μ, neither of which is observable to the firm.* The probability distribution of θ and ϵ is known to both firms and workers.

Most of the analysis focuses on a special case in which

$$Q_i = \mu_i\theta + \epsilon_i, \tag{1a}$$

where G is the distribution function of ϵ_i, g its density, and $E[\epsilon_i] = 0$, $E[\epsilon_i\epsilon_j] = 0$, and $E[\theta] = 1$. The linear equation for output is chosen for its analytical simplicity.[1] The effect of relaxing this assumption is discussed in Section 13. In relation to the principal-agent literature, equation (1a) is particularly important since Mirrlees' (1975) optimal income tax model can be viewed as a special case with $\sigma_\epsilon^2 = 0$ and Varian's (1980) analysis of social insurance can be viewed as a special case in which $\sigma_\theta^2 = 0$.

Our analysis will focus on a competitive market in which firms are risk neutral and expected profits are driven to zero. Compensation schemes in which a worker's expected payment is equal to his expected output will be called *feasible.* In the case of several agents, this assumes that the value of outputs $\{Q_1, \ldots, Q_n\}$ is just $\Sigma\, Q_i$. In a patent race, where $1/Q_i$ could be the discovery time for process i, it would be more reasonable to assume that only the max $\{Q_i\}$ matters and the definition of feasible would be different (Nalebuff and Varian, 1981).

Agents are identical. Their utility function is assumed to be additively separable in income, Y, and labor, μ. When income is a random variable, the expected utility, W, is

$$W = E[U(Y) - V(\mu(\theta))]. \tag{2}$$

The marginal utility of income is positive but weakly declining (individuals are not risk preferrers), while the disutility from effort is positive and increasing:

$$U'(Y) > 0, \qquad U''(Y) \leq 0, \qquad V'(\mu) > 0, \qquad \text{and} \qquad V''(\mu) > 0. \tag{3}$$

4. Sufficient statistics

■ Since only the set $\{Q_i\}$ is observable, compensation schemes are a function only of $\{Q_i\}$:

$$Y_i = Y(Q_1, \ldots, Q_i, \ldots, Q_n). \tag{4}$$

To achieve any greater flexibility, the principal must design a two-stage game involving signalling. With an appropriate incentive structure, individuals' announcements about θ can convey relevant information. These revelation schemes are discussed in Nalebuff and Stiglitz (1983).

A general interdependent compensation scheme is clearly no worse than an individualistic one,

$$Y_i = Y_i(Q_i). \tag{5}$$

The question is when is it better? Sometimes there will be no choice, and the principal will be forced to use a single output as the basis for all of the rewards. In the example

[1] This is not, of course, the most general form of linear structures: we could have written (1b) $Q_i = \mu_i\theta + \epsilon_i + \gamma + \mu_i\eta_i$, i.e., there is also an idiosyncratic effect on marginal productivity and a common effect on total output.

of a patent race, there may only be one observable output, the time to discovery. Those who lose redirect their efforts and generally do not produce any measurable output.

The advantage of using the multiple outputs to form the basis for each agent's compensation is that usually some information about θ can be gleaned from observing the whole array of Q_i's. This information can be used to make inferences about the μ_i's that permit the principal to tighten up and improve the reward scheme.

Theorem 1. When all the outputs are independent ($\sigma_\theta^2 = 0$), then observing all of the Q_i's provides no additional information about any agent's effort level, μ_i. In this case, the optimal compensation scheme is individualistic.

Proof. There is a simple constructive proof. Choose $Y_i(Q_i)$ such that

$$U(Y_i(\tilde{Q})) = E[U(Y_i(Q_1, \ldots, Q_i \ldots, Q_n | Q_i = \tilde{Q}))]. \tag{6}$$

A worker is given the same expected utility as a function of his output. Since he is risk averse, $Y_i(\tilde{Q}) < E[Y(Q_1, \ldots, Q_i, \ldots, Q_n | Q_i = \tilde{Q})]$. The principal can provide exactly the same incentives and expect to make a larger profit. The advantage of an individualistic reward scheme when agents are independent is a special case of a sufficient statistic theorem. Often it is possible to find a sufficient statistic, $T(Q)$, for the information about the common element of the agents' projects, θ. The optimal compensation scheme may then be simplified to

$$Y_i = Y_i(Q_i, T(Q)). \tag{7}$$

Intuitively, all of the information relevant to person i is contained in Q_i and $T(Q)$. The use of any other variable would only add random noise to the compensation and would thus be suboptimal.[2] A special application of the sufficient statistic theorem is the previous result, Theorem 1. When $\sigma_\theta^2 = 0$, then a sufficient statistic for θ is a constant, and, therefore, the optimal compensation scheme is individualistic. Other applications are also immediate. If both θ and ϵ are normally distributed random variables (with a known variance), then the average value of Q, $\bar{Q}$, is a sufficient statistic for $\theta\mu(\theta)$ and hence for θ. In the example of a classroom, when there is a large number of students, the average test score will reveal how hard the test was. The estimate of θ derived from $\bar{Q} = \theta\mu(\theta)$ will converge to the true θ as the number of contestants becomes large. In the limit we shall have a sufficient statistic for θ since we shall actually know θ. We can then apply the sufficient statistic theorem to show that the optimal second-best compensation scheme will depend only on an individual's output and θ, $Y_i = Y_i(\theta, Q_i)$. Even when the number of workers is limited, it may be advantageous to base compensation on the size of output relative to the mean, as this can reduce the noise associated with θ (as is done in the relative performance scheme in Section 11).

One can view the problem we are posing, of the choice of a payment scheme, as a statistical problem. The firm does not know the individual's level of effort. It attempts, from knowledge of the structure of the situation and observations of individual behavior, to make the "best" estimate, and then bases compensation on this estimated value. The difficulty with this approach is that the statistic used has both risk and incentive effects, and these must be taken into account in the choice of the "best" statistic.

5. The first-best optimum: a benchmark

■ In the subsequent discussion, it will prove useful to compare the results of alternative compensation schemes with the equilibrium with perfect information in which both μ and θ are observable. In the first-best optimum with risk neutral firms and risk averse

[2] We are indebted to Joe Farrell for extensive discussion on this question. For similar sufficient statistic results and proofs, see Farrell (1981), Grossman and Hart (1980), Holmstrom (1982), and Arrow (1970).

workers, the workers are given perfect insurance and receive a prespecified reward, $\bar{Y}$, that is dependent only on the expected output. Effort is supplied until the marginal utility of income multiplied by the increase in output with effort is just equal to the marginal disutility of work:

$$\theta U'(\bar{Y}) = V'(\mu(\theta)) \rightarrow \mu^*(\theta) \equiv V'^{-1}(\theta U'(\bar{Y})). \tag{8}$$

Hence

$$\bar{Y} = E[\theta\mu(\theta)] = E[\theta V'^{-1}(\theta U'(\bar{Y}))]. \tag{9}$$

6. The basic structure of contests

■ A rank order tournament is a compensation scheme in which contestants' rewards are based on their ordinal positions alone and not on the actual size of their output. With two contestants, there is a winner's salary, Y_w, and a loser's salary, Y_L. Unlike the standard marginal analysis, the winner's output is not necessarily worth Y_w. The winner is paid more than his marginal product so as to introduce a carrot to motivate greater effort among the contestants.

A principal maximizes his workers' utilities while being constrained by the competitive market to make zero expected profits. Hence, $Y_w + Y_L$ must equal the sum of the expected outputs, $2\bar{Q}$. (This is the dual of the problem in which the principal maximizes profits subject to the agents' utility constraints.) It is thus instructive to change notation. Let

$$\bar{Y} = E[Q_i] = \frac{Y_w + Y_L}{2} \qquad \text{and} \qquad x = \frac{Y_w - Y_L}{2}. \tag{10}$$

Any two prizes can always be thought of as a safe income, $\bar{Y}$, plus risk (or prize), x.

□ **Individual effort.** The individual's expected utility is a function of the probability of his winning. This in turn depends on his level of effort, his opponent's level of effort, the environmental variable θ, and the distribution of ϵ. For a given distribution of ϵ, the probability of winning is denoted by $P(\mu_1, \mu_2, \theta)$. An agent's expected utility is

$$W = PU(\bar{Y} + x) + (1 - P)U(\bar{Y} - x) - E[V(\mu)]. \tag{11}$$

By paying individuals on the basis of a contest, rather than on the basis of output, we take a random disturbance, ϵ, which has zero mean and is uncorrelated with effort, μ, and replace it by a disturbance, x, that is correlated with effort. If a contestant works harder, he is more likely to win, and thus the chance (P) of a positive x rises. In the symmetric equilibrium, the chance of winning is 1/2, so that the expectation of the prize (x or $-x$) is zero.

Workers will supply effort until their marginal disutility from work is just balanced by their increased chance of winning the value of the prize. After observing θ, effort is set so that

$$[\partial P(\mu_1, \mu_2, \theta)/\partial\mu_1]\Delta U - V'(\mu(\theta)) = 0, \tag{12}$$

where

$$\Delta U = U(\bar{Y} + x) - U(\bar{Y} - x).$$

This equation can be solved for μ_1 as a function of μ_2, and a symmetric reaction function giving μ_2 as a function of μ_1 can similarly be derived.[3] Although it should be apparent

[3] If μ_2 is a continuous, increasing function of μ_1 and if for large enough values of μ, $d\mu_2/d\mu_1 < 1$ (if "one" works harder so does "two," but his increment in work effort is smaller), then there exists a (symmetric) equilibrium. It is conceivable, however, that "two" decreases work effort in response to an increase by "one," and indeed, the response may be discontinuous. In this case there may exist no symmetric equilibrium, but there may be an asymmetric one, or no pure strategy equilibrium, or multiple equilibria (see discussion in the Appendix).

that it is possible that there exist asymmetric equilibria, with one individual working hard and the other slacking off, for this part we focus on the symmetric equilibrium with $\mu_1 = \mu_2$.

□ **The return to effort.** With two contestants, the agent with the greater output wins. If agent 1, working at effort level μ_1, is to beat agent 2, working at effort μ_2, it must be the case that the realizations of epsilon satisfy

$$\theta\mu_1 + \epsilon_1 > \theta\mu_2 + \epsilon_2. \tag{13}$$

The probability of this occurring for a given ϵ_2 is

$$1 - G(\theta(\mu_2 - \mu_1) + \epsilon_2). \tag{14}$$

To calculate the total probability of winning, we integrate over all possible values of ϵ_2, weighted by the density of ϵ_2, $g(\epsilon_2)$. Hence,

$$P(\mu_1, \mu_2, \theta) = \int [1 - G(\theta(\mu_2 - \mu_1) + \epsilon_2)]g(\epsilon_2)d\epsilon_2. \tag{15}$$

At the symmetric solution, $\mu_1 = \mu_2$, $P = \frac{1}{2}$, and player 1's increased chance of winning by working harder is

$$\partial P(\mu, \mu, \theta)/\partial\mu_1 = \theta \int g(\epsilon_2)g(\epsilon_2)d\epsilon_2 = \theta\bar{g}, \tag{16}$$

where $\bar{g} = E[g(\epsilon)]$.[4]

□ **The cornerstone equation of contests.** Substituting (16) into the first-order condition (12) yields the fundamental equation describing the equilibrium in contests:

$$\theta\Delta U\bar{g} = V'(\mu(\theta)). \tag{17}$$

The marginal disutility of labor just equals the utility value of the increased chance of winning the prize. For the contest to be feasible (zero expected profits), the expected compensation must equal expected output,

$$\bar{Y} = E[\theta\mu(\theta)] = E[\theta V'^{-1}(\theta\Delta U\bar{g})]. \tag{18}$$

The optimal contest is the pair $(\bar{Y}, x)$ that maximizes the contestants' expected utilities subject to the feasibility constraint (18).

□ **Central properties of contests.** The cornerstone equation illustrates one essential property of tournaments: effort varies with θ, while expected income does not. This adaptability to varied environmental states is one of the important characteristics of a good compensation scheme. It is even possible to replicate the first-best pattern of effort supply by setting

$$U'(\bar{Y}) = \bar{g}\Delta U, \tag{19}$$

as this implies

$$\theta U'(\bar{Y}) = \theta\Delta U\bar{g} = V'(\mu^*(\theta)). \tag{20}$$

Although it is possible to replicate the first-best level of effort, the contest will not at the same time yield the first-best level of utility. In tournaments, agents must bear risk because the prize is x or $-x$. Of course, *if individuals are risk neutral, then this variation*

[4] We may now see that the second-order conditions will be satisfied under a broad range of conditions. Differentiating (12), we require $\partial^2 W/\partial\mu^2 = \theta\Delta U \int g(\epsilon)g'(\epsilon)d\epsilon - V''(\mu) < 0$. If the disturbance is symmetric, then the first term drops out while the second term is negative by assumption. More generally, the second-order conditions will be satisfied if $\theta\Delta U[g^2(\bar{\epsilon}) - g^2(\underline{\epsilon})] < 2V''(\mu)$, where $\bar{\epsilon}$ and $\underline{\epsilon}$ are the maximum and minimum values on the support of ϵ.

is unimportant, and setting the prize according to (19) will achieve the full information first-best outcome. In general, the principal will choose to sacrifice some efficiency to reduce the risk borne by agents.

Theorem 2. A larger prize motivates greater effort and thus increases mean income.

Proof. To determine how effort varies with the prize, differentiate the worker's first-order condition (17),

$$\theta\bar{g}[U'(\bar{Y}+x)\{d\bar{Y}/dx+1\}-U'(\bar{Y}-x)\{d\bar{Y}/dx-1\})=V''(\mu)d\mu(\theta)/dx. \tag{21}$$

Multiplying through by θ and taking expectations,

$$d\bar{Y}/dx=\{\bar{g}SE[\theta^2/V''(\mu(\theta))]\}/\{1-\Delta U'\bar{g}E[\theta^2/V''(\mu(\theta))]\}>0, \tag{22}$$

where

$$S=U'(\bar{Y}+x)+U'(\bar{Y}-x) \qquad \text{and}$$

$$\Delta U'=U'(\bar{Y}+x)-U'(\bar{Y}-x)<0.$$

Although agents believe that by working harder they will increase their chance of winning, P, in fact both agents work equally harder, thereby leaving P at ½ in equilibrium. But as a result of their greater effort supply, the new feasible compensation scheme must commensurately increase $\bar{Y}$.

When the prize is sufficiently large to motivate the first-best level of effort, agents may find the losing state of nature particularly unpleasant. In such circumstances, $d\bar{Y}/dx$ can be greater than one and a contest has some of the features of a "rat race" (Akerlof, 1976). As the prize is increased, an agent works so much harder that his income in the losing state, $\bar{Y}-x+(d\bar{Y}/dx-1)dx$, is actually bigger than before; his increased effort more than compensates for the increase in the prize.

□ **The optimal contest.** We are now in a position to solve for the optimal prize, x, in a two-person rank order tournament. Recall that at the symmetric equilibrium, $P=$ ½ and effort supply is determined by (17), so that expected utility from (11) is

$$W=(1/2)[U(\bar{Y}+x)+U(\bar{Y}-x)]-E[V(V'^{-1}(\theta\Delta U\bar{g}))]. \tag{23}$$

Differentiating with respect to x, at the optimal prize, $\tilde{x}$,

$$\begin{aligned} dW/dx &= (1/2)S\bar{Y}'+(1/2)\Delta U'-E[V'd\mu/dx] \\ &= (1/2)S\bar{Y}'+(1/2)\Delta U'-E[\bar{g}\Delta U\theta d\mu/dx] \\ &= [(1/2)S-\bar{g}\Delta U]\bar{Y}'+(1/2)\Delta U'=0, \end{aligned} \tag{24}$$

where $\bar{Y}'$ is as in (22) and the value of $V'=\theta\Delta U\bar{g}$ is taken from (17).

Theorem 3. There is an undersupply of effort in the second-best solution. Observable increases in the agents' effort levels improve their welfare.

Proof. At the optimal prize, if the principal could observe a marginal increase in effort, the change in welfare would be

$$dW/d\mu=[(1/2)S-\bar{g}\Delta U]\bar{Y}'>0, \tag{25}$$

which follows directly from (24).

7. Properties of the second-best solution using contests

■ This section considers when a prize system is likely to be an effective compensation scheme.

□ **Problem from a nonconvexity.** Our model of a prize system will in general have a pure strategy Nash equilibrium. There is one important exception.

Theorem 4. When the variance of ϵ becomes small, there is a nonconvexity in the maximand that disrupts the second-best pure strategy solution.

Proof. As σ_ϵ^2 approaches zero, the increased chance of winning by working harder, $\theta\bar{g}$, must become infinite. By definition, $\bar{g} = E[g(\epsilon)]$. When the variance of a mean zero distribution approaches zero, the density collapses to a peak at the origin. The mean value of the density then becomes arbitrarily large. For example, when $g(\epsilon)$ is the normal density with mean 0 and variance σ_ϵ^2, then $\bar{g} = 1/[4\pi\sigma_\epsilon^2]^{1/2}$.

Intuitively, when there is a small amount of noise, a little more effort guarantees the contestant first prize. As $\bar{g}$ becomes large, the first-best level of effort can be generated with arbitrarily small prizes. But, when the prize is very small, the symmetric solution is only a local optimum. Why work at all? An agent who chooses not to supply effort μ^* and instead does *no* work is sure to lose. The prize is so small, however, that while he is forfeiting his chance at winning the vanishing amount ΔU, he saves all his disutility of effort, $V(\mu^*) - V(0)$.

Doing no work is not dropping out of the game. It is more akin to cheating on the contract. With the variance of the noise becoming arbitrarily small, there would be no doubt that this low output resulted from a lack of effort rather than bad luck. But the prize reward system is too simple to deter this violation. The rewards are based only on *ordinal* rank and so in this case, a mile is as good as a miss! Were different punishments established for different low levels of output, this would be a return to general nonlinear compensation schemes. Various resolutions of the problems presented by this nonconvexity are discussed in the conclusion, in Stiglitz (1980), and in Nalebuff (1982, chapt. 2). In the following analysis we must obviously restrict our attention to situations in which there is sufficient noise that the symmetric solution is indeed an equilibrium.

□ **Comparison with the first-best level of effort.** The claim in the summary that it is possible for the second-best optimum with a prize scheme to motivate more effort than in the first-best solution should be greeted with skepticism. If the first best is unobtainable because the principal cannot monitor effort, and effort is undersupplied, then it would initially seem doubtful that with a prize system we might desire more than the first-best optimal amount of effort. But the introduction of a prize forces the agent to bear some risk. The cost associated with this risk is related to his income. Agents with declining absolute risk aversion can reduce this cost by working harder, earning a higher income, and thus lowering their risk aversion. Examples in Nalebuff (1982) show that this effect can be sufficiently strong for the second-best solution to involve more effort supply than the first-best optimum.

This result is sensitive to the specification of the utility function. In the framework of Lazear and Rosen (1981), expected utility is defined by[5]

$$\tilde{W} = E[U(Y - V(\mu))]. \tag{26}$$

Here it must always be the case that the second-best level of effort is less than the first best. This result follows from the fact that the optimal supply of effort, $V'(\mu^*) = \theta$, is unaffected by any random fluctuations in income.

□ **The effect of changes in the variance in θ in a contest.** If the prize system is chosen optimally, changes in the variance of θ affect expected utility directly and through the

[5] One implication of this specification is that disutility of work is negatively correlated with income.

effect on effort (but not through an effect on x). The direct effect of the change in the distribution of θ is difficult to sign and in general will depend on whether the disutility of effort, $V(\mu(\theta))$, is concave or convex as a function of θ.

Theorem 5. With a quadratic disutility of effort, welfare is an increasing function of the variance of θ.

Proof.

$$V(\mu) = \mu^2/2 \Rightarrow \mu(\theta) = \theta\bar{g}\Delta U; \qquad \bar{Y} = \bar{g}\Delta U E[\theta^2]; \qquad E[V(\mu(\theta))] = (1/2)\bar{g}\Delta U\bar{Y}, \quad (27)$$

and

$$\partial\bar{Y}/\partial E[\theta^2] = \bar{g}\Delta U/\{1 - \bar{g}\Delta U' E[\theta^2]\} > 0. \quad (27a)$$

To calculate the total effect of a change in the distribution of θ, note that

$$W = (1/2)[U(\bar{Y} + x) + U(\bar{Y} - x)] - (1/2)\bar{g}\Delta U\bar{Y} \quad (28)$$

$$dW/dE[\theta^2] = \frac{dW}{dx}\frac{dx}{dE[\theta^2]} + \frac{\partial W}{\partial\bar{Y}}\frac{\partial\bar{Y}}{\partial E[\theta^2]} \quad (29)$$

$$= [(1/2)S - (1/2)\bar{g}\Delta U - (1/2)\bar{g}\Delta U'\bar{Y}]\partial\bar{Y}/\partial E[\theta^2] \quad (29a)$$

$$> [(1/2)S - \bar{g}\Delta U]\{\partial\bar{Y}/\partial E[\theta^2]\} > 0, \quad (29b)$$

where $dW/dx = 0$ and the final inequality both follow from the first-order condition (24) determining $\tilde{x}$. In this example, mean preserving increases in the spread of θ raise welfare. Hence, *the loss in welfare for zero variance in θ represents an upper bound to the loss in welfare from a contest.* More generally, this result will depend on assumptions concerning V''' which determine how the new distribution of effort affects output and disutility of work.

Stiglitz (1980) demonstrates conditions in which increases in the variance of θ improve welfare in a contest while they diminish welfare in a piece rate scheme. Intuitively, when θ is highly volatile, to give piece rate workers a sufficient incentive to supply effort, the piece rate must be positive and workers will bear the risk associated with large fluctuations in Q_i. In contrast, a contest can replicate the first-best level of effort and thus the monetary value of the loss in welfare relative to the first best is strictly less than the loss from the risk associated with the prize (which is of the order of $-x^2U''(\bar{Y})/2U'(\bar{Y})$) which stays relatively constant as σ_θ^2 increases.[6]

Although risk is normally associated with the variance of θ, it may also be appropriate to consider the effects of changes in the higher moments of θ. Consider a mean preserving spread that increases the variance of θ^2 while holding both the mean and variance of θ constant. In the example of quadratic disutility of effort, from (27) and (28) welfare is only a function of $E[\theta^2]$, and, hence,

$$dW/dE[\theta^4] = 0. \quad (30)$$

When workers have a quadratic disutility of effort, a contest is sensitive to the distribution of θ only through its first two moments.

8. Improving on the standard tournament

■ This section introduces two modifications of the standard tournament which improve the equilibrium outcome.

□ **Winning by a gap.** The idea behind any incentive scheme in the principal-agent problem is to provide large marginal incentives to work hard without exposing the agent

[6] This argument can be further generalized to show the conditions under which a contest will dominate even nonlinear piece rate schemes as the variance of θ becomes large.

to risk. The introduction of "gaps" can lower the probability that any prize will be paid while maintaining the same level of marginal incentives. Instead of ranking agents solely on the basis of the relative position of their outcomes, a principal can decide to rank one agent above another only if that agent's output is greater than his opponent's by a positive gap. In horse racing and Olympic swimming, first place can be determined by a head or a hair. But, in the marketplace, a firm is the leader in its industry only when it has a significantly higher market share than its competitors'.

In a contest with a gap, γ, there are three possible outcomes: win, lose, and tie. Let the reward for tying be $\bar{Y}$ and the gain for winning over losing be as before, ΔU. In the two-person contest, expected utility in the symmetric equilibrium is

$$W = (1 - 2P)U(\bar{Y}) + PU(Y + x) + PU(\bar{Y} - x) - E[V(\mu)]. \tag{31}$$

Agent 1 wins if

$$\theta\mu_1 + \epsilon_1 > \theta\mu_2 + \epsilon_2 + \gamma. \tag{32}$$

This leads to a transformation of equations (15) and (16):

$$P(\mu_1, \mu_2, \theta, \gamma) = \int [1 - G(\theta(\mu_2 - \mu_1) + \epsilon_2 + \gamma)]g(\epsilon_2)d\epsilon_2. \tag{15a}$$

At the symmetric solution, player 1's increased chance of winning by working harder is

$$\partial P(\mu, \mu, \theta)/\partial\mu_1 = \theta \int g(\epsilon_2 + \gamma)g(\epsilon_2)d\epsilon_2 = \theta\bar{g}(\gamma), \tag{16a}$$

where $\bar{g}(\gamma) = E[g(\epsilon + \gamma)]$. An increase in γ always lowers the probability that either agent will win and it lowers that probability by an amount just equal to $1/\theta$ times an agent's increased chance of winning by working harder,

$$\partial P/\partial\gamma = -\bar{g}(\gamma). \tag{33}$$

Theorem 6. If the distribution of the disturbance term is symmetric, or more generally if $g(\bar{\epsilon}) \geq g(\underline{\epsilon})$, where $\bar{\epsilon}$ and $\underline{\epsilon}$ are the maximum and minimum values on the support of ϵ, then introducing a strictly positive gap will improve the outcome of a tournament.

Proof. The proof proceeds by solving for the optimal gap, γ, and then showing that under the conditions of Theorem 6, the optimal γ is strictly positive. At the symmetric equilibrium,

$$\partial\bar{g}(\gamma)/\partial\gamma = E[g'(\epsilon + \gamma)]. \tag{34}$$

A change in γ affects an agent's effort supply and his expected level of output,

$$\partial\bar{Y}/\partial\gamma = \bar{g}'\Delta UE[\theta^2/V'']/\{1 - \bar{g}\Delta U'E[\theta^2/V'']\}. \tag{35}$$

The sign of this change is determined by the sign of $\bar{g}'$. As the prize, x, is chosen optimally, an increase in γ changes expected utility only through its effect on effort, expected output, and the probability of winning,

$$\partial W/\partial\gamma = -\bar{g}(\gamma)[U(\bar{Y} + x) + U(\bar{Y} - x) - 2U(\bar{Y})] + [PS + (1 - 2P)U'(\bar{Y}) - \bar{g}\Delta U]\bar{Y}'. \tag{36}$$

The first term is always positive by the concavity of $U(Y)$. At $\gamma = 0$, $P = \frac{1}{2}$, the second term will have the same sign as $\bar{Y}'$ since $(1/2)S > \bar{g}\Delta U$ from (24). Under the conditions of Theorem 6, $\bar{Y}'$ will also be positive,

$$\partial\bar{g}(0)/\partial\gamma = \tfrac{1}{2}[g^2(\bar{\epsilon}) - g^2(\underline{\epsilon})] > 0 \rightarrow \bar{Y}' > 0. \tag{37}$$

Hence, at $\gamma = 0$, $\partial W/\partial\gamma > 0$, and it is worthwhile to introduce a gap. More generally, to solve for the optimal gap requires finding the solution to $\partial W/\partial\gamma = 0$ from (36).

Implementing the use of gaps in competitive reward systems within an organization may also improve the sustainability of a contest. Workers who win contests and rise to the top of the managerial hierarchy have the power to redesign the incentive scheme. These top managers have had, on average, better than average luck. Many of the luckier ones know that were they subjected to repeated evaluation through a contest, they might lose their position. Thus, for a contest to be sustainable it is important to make skill and effort more influential than luck in determining winners. One solution to this problem is found in the asymmetric equilibrium with gaps (Nalebuff, 1982). If some agents work harder than others, winners are determined mostly by extra effort and only sometimes by luck.

□ **Contests with a large number of players.** Increasing the number of players does two things: the amount of information conveyed by output is expanded (as seen in the sufficient statistic result in Section 4), and the scope for designing reward schemes is increased. We start with a very simple generalization, a single prize to the one winner and postpone the discussion of tournaments with n prizes until Section 12.

When there are n agents competing for a single prize, we can think of each of them putting an amount x into a kitty and the winner collecting nx. At the symmetric equilibrium, expected utility is

$$W = U(\bar{Y} - x) + (1/n)\Delta U(n) - E[V(\mu)], \tag{38}$$

where $\Delta U(n) = U(\bar{Y} - x + nx) - U(\bar{Y} - x)$.

As before, we can calculate the expected return to increased effort. Assume that everyone else supplies effort $\bar{\mu}$, and the ith individual supplies effort μ_i; then, if he obtains a realization ϵ_i, he wins only if

$$\epsilon_j \leq \epsilon_i + \theta(\mu_i - \bar{\mu}) \qquad \text{for all } j. \tag{39}$$

The chance of this occurring is $G(\epsilon_i + \theta(\mu_i - \bar{\mu}))^{n-1}$. The probability of winning is the expectation of this chance conditional on ϵ_i,

$$P = \int G(\epsilon_i + \theta(\mu_i - \bar{\mu}))^{n-1} g(\epsilon_i) d\epsilon_i. \tag{40}$$

At the symmetric equilibrium, $P = 1/n$, and the marginal increase in the chance of winning by working harder, $\partial P/\partial \mu$, is

$$\partial P/\partial \mu = \theta \bar{g}(n) = \theta(n-1) \int [G(\epsilon)]^{n-2} g^2(\epsilon) d\epsilon. \tag{41}$$

The principal can determine a prize, $x^*(n)$, such that each agent will supply the same effort as in the first-best optimum by setting

$$\bar{g}(n)\Delta U(n) = U'(\bar{Y}). \tag{42}$$

In rewriting (19) as (42), we note that provided $\bar{g}(n)$ goes to zero slower than $(1/n)$, then $\Delta U(n)/n$ must in the limit be zero,

$$\lim_{n\to\infty} n\bar{g}(n) \cdot (\Delta U(n)/n) = U'(\bar{Y}) \Rightarrow [\lim_{n\to\infty} n\bar{g}(n) = \infty \rightarrow \lim_{n\to\infty} (\Delta U/n) = 0]. \tag{43}$$

Recalling the expected utility from (11), if the utility function has a linear asymptote[7] (so that marginal utility is bounded below) and $\Delta U(n)/n$ approaches zero, then the prize x must also approach zero and the level of utility will tend to the first best:

$$\lim_{n\to\infty} \Delta U(n)/n > nxU'(\bar{Y} + nx)/n = xU'(\bar{Y} + nx) = 0 \rightarrow \lim_{n\to\infty} x^*(n) = 0, \tag{44}$$

[7] Otherwise, if $U'(Y)$ tends to zero, then for any finite x, $\Delta U(n)/n$ approaches zero, and the nonconvexity problem will still arise.

$$\lim_{n\to\infty} W = U(\bar{Y}) - E[V(\mu^*)], \text{ the first-best level of utility.} \tag{45}$$

But, *the symmetric equilibrium is disrupted by nonconvexity constraints.* An agent who chooses to do no work only loses ΔU with probability $(1/n)$ while saving disutility of effort $V(0) - V(\mu^*)$. The agent who chooses to do no work has an expected utility

$$\lim_{n\to\infty} W_{|\mu_i=0} = U(\bar{Y}) - V(0) > U(\bar{Y}) - E[V(\mu^*(\theta))]. \tag{46}$$

Increasing the size of the tournament was initially attractive as it provided better incentives for smaller risks. Unfortunately, it also changes the rewards in a way that makes breaking the equilibrium more attractive.

□ **Penalties.** The solution to the nonconvexity problem is suggested in Mirrlees' (1974) discussion of a punishment scheme for workers who fail to meet their quota. In the framework of contests, this implies a penalty for the single worker with the lowest output. As the number of agents increases, the chance of being the one with the lowest output becomes very small. But it is the marginal conditions that are important in determining the size of the penalty. Provided the marginal incentives do not fall too quickly, the principal will not need to give increasingly horrific punishments to the loser.

Mathematically, the punishment model is identical to the prize model if we replace x by $-x$ and the chance of winning by the chance of losing. At the symmetric equilibrium,

$$W(n) = U(\bar{Y} + x) - \Delta U(n)/n - E[V(\mu(\theta))], \tag{47}$$

where $\Delta U(n) = U(\bar{Y} + x) - U(\bar{Y} + x - nx)$. As before with a single prize, the first-order condition determining effort is

$$-\theta \Delta U(n)\tilde{g}(n) - V'(\mu(\theta)) = 0, \tag{48}$$

where $\theta\tilde{g}(n)$ is the marginal *decrease* in the chance of losing by working harder, and

$$\tilde{g}(n) = -(n-1)\int [1 - G(\epsilon)]^{n-2} g^2(\epsilon) d\epsilon. \tag{49}$$

In general, when the error distribution is on a support $[-\infty, \infty]$, the marginal change in the probability of losing, $\theta\tilde{g}$, will tend to zero as the number of players becomes very large. But, provided the approach is slower than $1/n$ (that is, the limit as $n \to \infty$ of $n\tilde{g}(n) = -\infty$), then the expected utility of the agents will approach the first-best optimum and a pure strategy Nash equilibrium will exist.

To replicate the first-best level of effort supply, choose x such that

$$U'(\bar{Y}) = -\tilde{g}(n)\Delta U = -n\tilde{g}(n)[\Delta U(n)/n]. \tag{50}$$

As $-n\tilde{g}$ approaches infinity, the prize must be such that $\Delta U(n)/n$ tends to zero. If $\Delta U(n)/n$ goes to zero, then the prize must also approach zero,

$$\Delta U(n)/n > nxU'(\bar{Y} + x)/n = xU'(\bar{Y} + x). \tag{51}$$

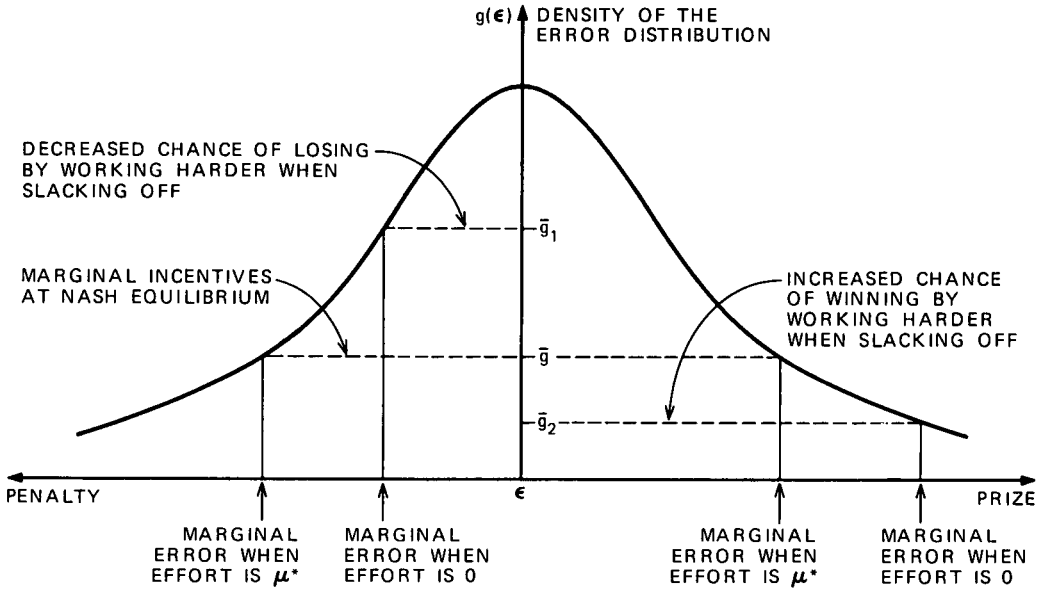
Thus, welfare approaches the first best,

$$W = \lim_{n\to\infty} U(\bar{Y} + x) - \Delta U(n)/n - E[V(\mu^*(\theta))] = U(\bar{Y}) - E[V(\mu^*(\theta))]. \tag{52}$$

In the penalty framework, the solution with all workers supplying effort levels $\mu^*(\theta)$ will be a Nash equilibrium. In the model with only a single winner, shirking costs at most $\Delta U(n)/n$, which falls with n eventually to zero. Here, shirking may cost as much as $(n-1)\Delta U(n)/n$, which rises with n up to $\Delta U(n)$. Shirking may not cause a player to forfeit his entire chance of winning. The agent's maximization problem is concave because (i) the more an agent shirks, the less he saves on the margin as his marginal disutility of

FIGURE 1

CHEATING WHEN THERE IS A PENALTY SHIFTS THE MARGINAL EPSILONS TOWARDS 0 AND A REGION WITH A HIGHER DENSITY. CHEATING WHEN THERE IS A PRIZE SHIFTS THE MARGINAL EPSILONS UPWARDS, BUT AWAY FROM THE ORIGIN, TOWARDS A REGION WITH A LOWER DENSITY.



effort is assumed to be increasing, (ii) the more an agent shirks, the marginal epsilons on which he loses will be smaller and thus more likely and hence the greater is his marginal cost in terms of an increased chance of losing. More formally, for unimodal distributions with sign $(g'(\epsilon)) = -\text{sign}(\epsilon)$, the marginal decrease in an agent's chance of losing becomes increasingly negative the more he shirks.

In a contest with just a prize to the winner, as an agent cheats more, the marginal epsilons on which he wins become greater and thus less likely so that the marginal cost from cheating falls and the problem has a nonconvexity. This is illustrated in Figure 1. There is a danger in trying to translate Mirrlees' argument suggesting infrequent but large penalties into high but unlikely bonuses. Holmstrom (1982, Theorem 4; 1977, p. 242) claimed "if the agent's utility function has a linear asymptote and the principal is risk neutral, then we can essentially reproduce the earlier argument to conclude that first-best approximations via high but unlikely bonuses are possible." The problem may be locally concave (so that the second-order conditions are locally satisfied), but if the first best is to be approximated, then the expected prize must go to zero. An agent who cheats can do no worse than lose the expected prize which is going to zero while saving all of his disutility of effort. The first-order conditions do not describe the global optimum. In contrast, with a penalty, the more an agent cheats, the more his costs rise. If he could cheat enough, the expected penalty would go from zero to a sure loss of $-\Delta U(n)$.

9. Individualistic schemes: piece rates and quotas

■ The two simplest and most commonly studied compensation schemes are the linear piece rate system and the quota system. Here we analyze these reward devices to compare them with contests.

The zero-profit condition implies that an agent's compensation under a linear piece rate scheme is

$$Y_i = \alpha Q_i + (1 - \alpha)\bar{Q}. \tag{53}$$

An individual receives as compensation the expected output plus an incentive bonus based on the difference between his output and the expected output of the group. The first-order condition determining effort supply[8]

$$\alpha\theta E[U'(\alpha Q_i + (1-\alpha)\bar{Q})] = V'(\mu(\theta)). \tag{54}$$

As in the contest, μ adjusts to θ. But the response is less than with a contest or in the first-best optimum,

$$\frac{d\mu}{d\theta} = \alpha[E(U' + \alpha\theta\mu U'')]/[V'' - \alpha^2\theta^2 EU'']. \tag{55}$$

It is even possible for μ to move inversely with θ. When σ_θ^2 and σ_ϵ^2 are small, then for logarithmic utility $d\mu/d\theta \approx 0$.

For a fixed piece rate and small σ_ϵ^2, a change in the variance of ϵ increases or decreases effort according to $U''' > 0$ or $U''' < 0$. In contrast, the effect in a contest from the change in the distribution of ϵ depends only on the effect on $\bar{g}$.

The simplest nonlinear individualistic payment scheme is the quota (or bonus) scheme, with

$$Y = \begin{cases} Y_1 & \text{if} \quad Q > \hat{Q} \\ Y_2 & \text{if} \quad Q < \hat{Q}. \end{cases} \tag{56}$$

The individual chooses μ to

$$\max\ U(Y_1)(1 - G(\hat{Q} - \mu\theta)) + U(Y_2)G(\hat{Q} - \mu\theta) - V(\mu), \tag{57}$$

which has a first-order condition

$$\theta\Delta\tilde{U}g(\hat{Q} - \mu\theta) = V'(\mu(\theta)), \tag{58}$$

where $\Delta\tilde{U} = U(Y_1) - U(Y_2)$.

There are two immediate difficulties with the use of such a simple quota or bonus scheme. First, even when the worker is risk neutral, the bonus scheme *cannot* replicate the first-best solution (because $g' \neq 0$ unless g is uniform). Second, there is a nonconvexity problem that arises when θ is small. When the quota is high and the task is difficult, then the worker may decide that it is not worth competing for the bonus.

10. Comparison of contests and piece rates

■ A prize system is likely to be better than a piece rate when the range of outputs is highly variable. A contest truncates the extreme possibilities, restricting the risk to $\pm x$. In a piece rate system, the worker must accept a small chance of a very small income.

Both the piece rate and the contest are able to achieve the first-best level of effort when agents are risk neutral. In the piece rate system, the principal sets the piece rate, α, equal to one. There is no loss from risk, and incentives are identical to what they would be in the first-best optimum. In the contest, the principal can induce the first-best level of effort state by state by setting the prize x according to (19), $x^* = 1/(2\bar{g})$. Having motivated $\mu^*(\theta)$, the prize does not affect the agent's utility because its expectation is zero and the agent is risk neutral.

Although there are many ways to introduce risk aversion, we consider just the simple example of a quadratic utility function so that

$$W = E\{a + bY - c(Y - Y^*)^2 - \mu^2/2\}, \tag{59}$$

where Y^* is the expected output in the first-best solution. Of course, $c = 0$ corresponds to risk neutrality. We start by considering changes in the expected utility of the contest as c is increased from 0,

[8] As discussed in Arnott and Stiglitz (1980), there may be problems with the second-order conditions.

$$dW^C/dc_{|c=0} = (dW^C/dx)(dx/dc) + (\partial W^C/\partial\mu)(\partial\mu/\partial c) + \partial W^C/\partial c. \tag{60}$$

The first term is zero as the prize, x, is chosen optimally. Effort supply is determined by

$$\mu = \theta\bar{g}\Delta U = \theta\bar{g}2bx. \tag{61}$$

The optimal effort level is not directly affected by changes in c when $c = 0$ and thus the second term is also zero. The utility loss from an increase in risk aversion is thus proportional to the variance of the risk, x. This result does not depend on either σ_θ^2 or σ_ϵ^2 being small. At $c = 0$, $x = 1/[2\bar{g}]$ and

$$dW^C/dc_{|c=0} = \partial W^C/\partial c = -E[(Y - Y^*)^2] = -x^2 = -1/[4\bar{g}^2]. \tag{62}$$

In the piece rate system,

$$dW^p/dc_{|c=0} = (dW^p/d\alpha)(d\alpha/dc) + (\partial W^p/\partial\mu)(\partial\mu/\partial c) + (\partial W^p/\partial c). \tag{63}$$

As α is optimally chosen at 1, a slight change in α will have no effect. From the worker's optimal choice of effort supply, the decrease in consumption due to less effort is just offset by the gain in leisure, $\partial W^p/\partial\mu = 0$. Again, the only term that is important is the direct effect of c. The variance of output is the sum of the variance of ϵ and the variance of $\theta\mu(\theta)$. At $c = 0$, $\mu = b\theta$, so that the variance of $\theta\mu(\theta)$ is b^2 Var (θ^2):

$$dW^p/dc_{|c=0} = (\partial W^p/\partial c) = -(\sigma_\epsilon^2 + b^2 \text{ Var } (\theta^2)). \tag{64}$$

Comparisons between the welfare of contests versus linear piece rates are very easy to make for small c,

$$W^C > W^p \qquad \text{if} \qquad (\sigma_\epsilon^2 + b^2 \text{ Var}(\theta^2) > 1/[4\bar{g}^2]. \tag{65}$$

Clearly, for large enough values of Var (θ^2) the contest will always be preferred. Even when Var $(\theta^2) = 0$, it is possible to find distributions of ϵ such that the contest is better (for example $g(\epsilon) = \frac{5}{2}\epsilon^4$ on $[-1, 1]$ and 0 elsewhere).[9]

11. Relative performance

■ The distinctive feature of contests is that only information about *rank* is used in determining compensation. We consider here some simple schemes in which the *magnitude* of relative performance enters into the compensation scheme. This will enable us to obtain some intuition concerning the kinds of situations where contests may be preferred.

Consider the compensation scheme

$$Y_i = \beta(Q_i - Q_j) + \bar{Y}. \tag{66}$$

The amount received by the individual is a fixed sum, $\bar{Y}$, plus a linear function of the difference between the two individuals' levels of output. In the symmetric equilibrium

$$Y_i = \beta(\epsilon_i - \epsilon_j) + \bar{Y}, \tag{67}$$

and hence the variance of the individual's income is just $2\beta^2\sigma_\epsilon^2$.

The two agents' outputs have identical correlation with θ, but independent ϵ_i's. Relative to an individualistic piece rate, using a second agent's output as a negatively correlated "asset" eliminates all the uncertainty in income due to the variance of $\theta\mu(\theta)$ at the expense of adding the variance of another ϵ. This scheme will be preferred to an individualistic piece rate if $E(\epsilon^2)$ is sufficiently smaller than $E[(\theta\mu(\theta))^2]$.

---

[9] From the sufficient statistic theorem we know that when θ is constant, the optimal nonlinear piece rate is superior to a contest. The fact that it is possible to demonstrate a prize system that dominates a linear piece rate when θ is constant is conclusive proof that a linear piece rate is far from optimal when the utility function is quadratic with only a slight amount of risk aversion.

This compensation scheme is further illuminated in the context of Holmstrom's (1982) sufficient statistic results. With a large number of agents, compensation can be based on the difference between an individual's production and the average output. As the average output converges to the true $\theta\mu(\theta)$, agents need only bear the risk associated with their individualistic noise, ϵ_i.

Compensation based on the magnitude of the differences between two agents' outputs can motivate the first-best level of effort by choosing β (potentially greater than 1) according to

$$\beta E[U'(\beta(\epsilon_1 - \epsilon_2) + \bar{Y})] = U'(\bar{Y}). \tag{68}$$

This is not the case, however, if we base compensation on the *ratio* of the individual's output to the mean output when there is a large number of agents, $Y_i = f(Q_i/\bar{Q})$,

$$\frac{Q_i}{\bar{Q}} = \frac{\mu_i}{\mu^*} + \frac{\epsilon_i}{\theta\mu^*}. \tag{69}$$

Individuals will set μ_i so that

$$\frac{1}{\mu^*} E[U'f'] = V'(\mu_i). \tag{70}$$

Here, μ varies with θ only through the effect of f. Thus,

$$\frac{d\mu}{d\theta} = \frac{-E\{(U''f'^2 + U'f'')\epsilon_i\}/\mu^2\theta^2}{V'' - E(U''f'^2 + U'f'')/\mu^2}. \tag{71}$$

If $\sigma_\epsilon^2 = 0$, μ is invariant to θ in contrast to the first-best solution, where an increase in the productivity of the individual (θ) leads to an increase in his effort. This relative performance criterion lacks the important property of "flexibility" we emphasized earlier.

12. Extensions: tournaments with multiple prizes

■ As the number of participants in a tournament becomes large, it is possible to improve on a single-winner or single-loser tournament by taking advantage of the increased flexibility of having n different prizes. Let the probability that an agent i finishes in position j up from the bottom be $P(Q_i = j)$. Then

$$P(Q_i = j) = \int \frac{(n-1)!}{(n-j)!(j-1)!} g(\epsilon)G\{\theta(\mu_i - \bar{\mu}) + \epsilon\}^{j-1}[1 - G\{\theta(\mu_i - \bar{\mu}) + \epsilon)\}]^{n-j}d\epsilon, \tag{72}$$

where $\bar{\mu}$ is the effort level of the other $n - 1$ agents. At the symmetric equilibrium ($\mu_i = \bar{\mu}$), agent i has an equal chance of ending up in any position, $P(Q_i = j) = 1/n$. When the prize for coming in at position j is $\tilde{\omega}_j$, then expected utility is

$$W = (1/n)[\sum_{j=1}^{n} U(\tilde{\omega}_j)] - E[V(\mu(\theta))]. \tag{73}$$

Effort supply is determined by

$$\sum_{j=1}^{n} \frac{\partial P(Q_i = j)}{\partial \mu_i} [U(\tilde{\omega}_j)] - V'(\mu(\theta)) = 0. \tag{74}$$

By working harder at the symmetric equilibrium there is an increased (or decreased) chance of coming in at position j,

$$\partial P(Q_i = j)/\partial \mu_i = \theta \int \frac{(n-1)!}{(n-j)!(j-1)!} g^2(\epsilon)(1 - G(\epsilon))^{n-j-1}G(\epsilon)^{j-2} \times \{(j-1)(1 - G(\epsilon)) - (n-j)G(\epsilon)\}d\epsilon. \tag{75}$$

It is tempting to hypothesize that a tournament with enough different positions can essentially replicate any general individualistic nonlinear compensation scheme. The idea

would be to divide the general nonlinear scheme into percentiles and then to approximate it by a reward structure that is flat over the range of a percentile (as is done in Green and Stokey (1981)). The reward in a contest for an agent who has rank j is then equalized with the payment to an agent who places in the $100j/n$th percentile. While these two compensation schemes yield the same expected utility at the symmetric solution, the marginal incentives are not the same. Thus, the equality cannot in general be sustained in a Nash equilibrium. In a contest, from equations (74) and (75), the optimal effort supply satisfies

$$V'(\mu(\theta))/\theta = \text{constant (independent of } \theta). \tag{76}$$

For the general nonlinear piece rate scheme, $Y(Q_i)$, the first-order condition determining effort supply does not have disutility of effort proportional to θ,

$$\theta E[U'\{Y(\theta\mu(\theta) + \epsilon)\}Y'(Q)] = V'(\mu(\theta)). \tag{77}$$

The nonlinear piece rate scheme is only able to duplicate a contest (and *vice versa*) when either (i) θ is not variable (Green and Stokey (1981) considered a model where the common disturbance did not affect the marginal productivity of effort), or (ii) the agent is induced to act as if he were risk neutral, $U(Y(Q)) = z_1 + z_2Q$ defines $Y(Q)$.

There are two special cases where it is relatively easy to consider the optimal prize structure in a rank order tournament with n positions,

(1) uniform density, $g(\epsilon) = \bar{g}\ \{\epsilon{:}[-A, A]\}$,
(2) exponential density, $g(\epsilon) = e^{\epsilon-1}\ \{\epsilon{:}[-\infty, 1]\}$.

With a uniform error distribution, prizes $2, \ldots, n-1$ contribute nothing towards work incentives. The chances of moving in and moving out of any interior position are the same. Hence, the net increase in the probability of winning any interior prize by working harder is zero. All the incentives must be provided by $\tilde{\omega}_1$ and $\tilde{\omega}_n$. To minimize risk, the principal will find it optimal to equalize all of the interior prizes, $\tilde{\omega}_2 = \tilde{\omega}_3 = \cdots = \tilde{\omega}_{n-1}$. Even with n potential prizes, the most general reward schemes with a uniform error distribution require only three different prizes: a last place, a first place, and a constant middle prize. This result has to be qualified to the extent that as the number of positions becomes large, the principal may have to use a greater number of penalties to ensure that the agents choose the interior maximum.

With an exponential density, it can be shown that increasing effort supply reduces *equally* the probability of getting anything but the top prize. Because the marginal incentives are exactly the same for $n-1$ of the prizes, these prizes should be equalized to minimize risk. Hence, the principal only needs two distinct prizes to achieve the general second-best solution using a tournament.

Although as before there is still the qualification about nonconvexities, these examples show that the study of tournaments with a simple reward structure may indeed approximate the solution for a more complicated rank order tournament.

13. Extension: when output is nonlinear in effort

■ The analysis up to this point has considered only cases in which output, Q_i, is linear with regard to effort, $Q_i = \theta\mu_i + \epsilon_i$. A more general framework is to have output as a function of effort. If we think of ϵ as a measurement error, then

$$\text{(i) } Q_i = F(\theta\mu_i) + \epsilon_i.$$

A second (and perhaps more interesting) case is when ϵ may be reasonably interpreted as a luck factor in which case

$$\text{(ii) } Q_i = F(\theta\mu_i + \epsilon_i).$$

Recall that effort is chosen once the common environmental factor, θ, is known.

In case (i), the first best is characterized by:

$$\bar{Y} = E[F(\theta\mu_i(\theta)] \tag{78}$$

and

$$\theta U'(\bar{Y})F'(\theta\mu_i(\theta)) = V'(\mu_i) \text{ defines } \mu_i(\theta). \tag{79}$$

In a tournament effort is supplied to increase one's chances of winning. Player 1 wins if $F(\mu_1\theta) - F(\mu_2\theta) + \epsilon_1 > \epsilon_2$. At the symmetric solution, $\mu_1 = \mu_2$, working harder increases the probability of winning by a factor proportional to F',

$$\partial P(\mu, \mu, \theta)/\partial\mu_i = P_{\mu_i} = \theta F'(\theta\mu_i(\theta)) \int g^2(\epsilon)d\epsilon = \theta F'(\theta\mu_i(\theta))\bar{g} \tag{80}$$

and

$$P_{\mu_i}\Delta U = V'(\mu_i(\theta)) \text{ defines } \mu_i(\theta). \tag{81}$$

The principal can *still* replicate the first-best level of effort by choosing a prize such that $\bar{g}\Delta U = U'(\bar{Y})$. The analysis of tournaments may be carried out as before.

It is more interesting to consider the interpretation where ϵ is like a luck factor. For this case, in the first best,

$$\theta U'(\bar{Y})\bar{F}'(\theta\mu_i(\theta)) = V'(\mu_i(\theta)) \text{ defines}^{10} \mu_i. \tag{82}$$

With a tournament in this situation, the presence of the production function F is irrelevant in determining the winner of the process as it is nothing more than a monotonic transformation of $\mu\theta + \epsilon$. That is,

$$F(\theta\mu_1(\theta) + \epsilon_1) > F(\theta\mu_2(\theta) + \epsilon_2) \Leftrightarrow \theta\mu_1(\theta) + \epsilon_1 > \theta\mu_2(\theta) + \epsilon_2. \tag{83}$$

Hence, the first-order condition determining work effort is just

$$\theta\bar{g}\Delta U = V'(\mu_i(\theta)) \text{ defines } \mu_i(\theta). \tag{84}$$

Unfortunately, the expression is now *missing* the term F' and thus except when F is linear, we can no longer replicate the first-best effort supply using a tournament. Hence, even when agents are risk neutral the first best is not achievable.

14. Prizes that influence choice of techniques

■ The choice of techniques when the reward system has a prize structure was initially studied in the context of credit rationing by Stiglitz and Weiss (1981). A loan from a bank is like a quota coupled with a linear piece rate. Should the agent earn less than the interest payment, he goes bankrupt and is rewarded zero. If he has produced more than his interest payment, the agent may keep the surplus. A rise in the interest rate (quota level) results in "riskier" (Rothschild and Stiglitz, 1970) strategies' being adopted by the borrowers.

Credit rationing may result from lending banks' being unwilling to accept the greater risks that accompany higher interest rates. But in other circumstances these greater risks would be welcomed. In contrast to the credit market, a potential advantage of competitive compensation schemes (i.e., contests or quotas) over piece rates is that they encourage entrepreneurs to disregard their natural risk aversion and choose more profitable although riskier strategies. This benefit is especially great in fields like research and development where prizes (in the form of patents) encourage risk taking that can dramatically shorten the time to discovery.

---

[10] The other variables are defined by $\bar{F} = E[F(\theta\mu(\theta) + \epsilon)]$, the expectation is taken over ϵ and $\bar{Y} = E[\bar{F}]$, the expectation is taken over θ.

The compensation schemes we have been considering simply do not have enough instruments to separately control both inputs (effort) and techniques.

15. Concluding remarks

■ Society may find the aftermath of competition, that is the presence of losers, to be unacceptable. An advantage of individualistic schemes is that *everyone* can have a high output, meet quota, and be rewarded. By contrast, when the principal does not know where to set the quota and thus bases rewards on relative performance, someone must come in last and be a loser. A problem with penalties is that they can become self-perpetuating. The loser becomes demoralized, fails to continue competing, and thus continues to lose. But prizes alone are ineffective in motivating effort from a large group as the expected prizes are too small to stop shirking.

Competition works best when all the participants are similar. A difficulty with the implementation of penalties is that the losers are usually more than just unlucky, they are often not so able. For them to compete, they would have to work harder than the average worker. Worse, the presence of a sure loser destroys everyone's incentives to work hard. When the different relative abilities are known, handicapping (i.e., as in golf tournaments) can restore the competitive environment that arises in a "fair" contest.

In noncompetitive production arrangements, considerable reliance may be placed on *social* incentives rather than on economic incentives. Individuals do not shirk because of the disapproval it generates. The nonconvexity problem with a small prize and a large number of players arises from each worker's desire to do no work and still reap the benefits of the loser's prize. Although the manager (principal) is assumed unable to monitor the agents' effort levels, it is likely that workers can observe each other. A worker may be reluctant to report a coworker who is doing only slightly less work, but the cheating strategy involves a discreet reduction in effort. If we can penalize cheaters rather than just giving them the loser's prize, then a pure strategy Nash equilibrium will exist even for the smallest of prizes. The possibility of obtaining the first-best solution gives each individual an incentive for extending disapproval to those who shirk (cf. Akerlof and Soskice's (1976) theory of sanctions).

The use of competitive compensation schemes seems less widespread than their evident advantages would suggest. This may be a result of important aspects of worker satisfaction which traditional economic models ignore. These considerations (i.e., work environment, group homogeneity) are probably less important in the analysis of competition between firms than in the competition *within* a firm.

In situations where contests are frequently observed, such as a patent race, there may be technological returns to cooperation, as in sharing information. In the competitive system, there are no incentives for cooperation. There are even rewards from engaging in destructive activity if it can hurt one's rival more than oneself. The piece rate system will encourage agents to cooperate when it is mutually beneficial, and this potentially may be very important. When the principal's reward function is strictly convex with respect to his agents' outputs, there is a further tradeoff in using contests. Although agents may have better incentives to work, there is an efficiency loss to the principal of having more than one project in operation (Nalebuff and Varian, 1981).

The models presented in this article have argued that there is a distinct role for competition—real competition, in the sense the word is ordinarily used, not the peculiar static sense in which much of neoclassical economics has come to use the term—in situations where there is imperfect information about the difficulties associated with different tasks, where it is prohibitively costly to observe inputs directly, and where it is difficult to measure the outputs with precision. The advantage of competitive systems is

that they have greater flexibility and greater adaptability to change in the environment than do other forms of compensation.

Appendix

Mixed strategy equilibria

■ We have shown in the text that in the simple prize to the winner contest, no symmetric pure strategy Nash equilibrium exists either as the number of contestants becomes large or as the variance (risk) due to epsilon vanishes. In both instances, the expected prize is sufficiently small that it is not worth competing for it. Thus, both agents would prefer to do no work and earn the loser's reward. This difficulty can be partially eliminated by combining performance standards with contests when ϵ has a finite support (Stiglitz, 1975). Here we consider the mixed strategy equilibrium.

Assume, as the extreme case, that $\sigma_\epsilon^2 = 0$. Clearly, any individual can, by increasing his level of effort slightly above that of his rival, assure his victory. Each contestant's reaction function is discontinuous, and no pure strategy equilibrium exists. Let $H(\mu)$ be the probability distribution of the maximum of the effort levels facing an individual competitor. In a mixed strategy equilibrium, the expected utility of an individual when pursuing any effort level, μ_i, must be constant,

$$U(\bar{Y} - x) + H(\mu_i)\Delta U - V(\mu_i) = k. \tag{A1}$$

Clearly, if

$$k = U(\bar{Y} - x) - V(0) \qquad \text{and} \tag{A2}$$

$$H(\mu) = [V(\mu) - V(0)]/\Delta U, \tag{A3}$$

the individual is indifferent as to his level of effort. Let the distribution of an individual's effort supply be $R(\mu)$. With n players, each agent faces $n - 1$ opponents,

$$H(\mu) = R(\mu)^{n-1} \tag{A4}$$

$$R(\mu) = \{[V(\mu) - V(0)]/\Delta U\}^{1/[n-1]}. \tag{A4a}$$

The expected output is

$$\bar{Y} = \int_0^{\tilde{\mu}} \mu r(\mu) d\mu = \int_0^{\tilde{\mu}} [1 - R(\mu)] d\mu, \tag{A5}$$

where the maximum level of effort $\tilde{\mu}$ satisfies

$$V(\tilde{\mu}) - V(0) = \Delta U. \tag{A6}$$

The optimal *feasible* mixed strategy equilibrium is determined by the choice of a prize, x^*, that maximizes expected utility, k. From equation (A2) this implies

$$U'(\bar{Y} - x)[\bar{Y}' - 1] = 0. \tag{A2a}$$

Thus, we shall look for the solution to $\bar{Y}' = 1$. From (A5) and (A6),

$$d\bar{Y}/dx = [(\tilde{\mu} - \bar{Y})S]/[(n - 1)\Delta U - (\tilde{\mu} - \bar{Y})\Delta U']. \tag{A7}$$

To provide a complete example of the mixed strategy solution, we further assume

(i) $V(\mu) = \mu^2/2$ and

(ii) $U(x) = x$, risk neutrality.

We choose a risk neutral utility function because it is analytically simple and because deficiencies of the mixed strategy solutions will be most apparent when compared with the first-best outcome that is achievable by using a simple piece rate. Substituting into (A5) and (A6) yields

$$\tilde{\mu}^2/2 = 2x, \tag{A6a}$$

$$\bar{Y} = \int_0^{\tilde{\mu}} \{1 - (\mu/\tilde{\mu})^{2/(n-1)}\} d\mu = 2\tilde{\mu}/(n+1). \tag{A5a}$$

At x^*, $d\bar{Y}/dx = 1$, which from equation (A7) implies

$$(n-1)x = [\tilde{\mu} - \bar{Y}]. \tag{A7a}$$

Thus at the optimal mixed strategy equilibrium

$$\tilde{\mu} = 4/(n+1), \qquad x^* = 4/[(n+1)^2], \qquad \text{and} \qquad \bar{Y} = 8/[(n+1)^2], \tag{A8}$$

$$\text{expected utility: } EU = U(\bar{Y} - x^*) - V(0) = \bar{Y} - x^* = 4/[(n+1)^2].$$

In the first-best solution, workers choose μ to maximize

$$EU = \mu - \mu^2/2 \rightarrow \mu^* = 1 \qquad \text{and} \qquad EU = .5 \tag{A9}$$

We observe that with only two workers, the mixed strategy is able to come reasonably close $\left(\frac{4}{9}\Big/\frac{1}{2}\right)$ to the first-best level of utility. But as the number of contestants increases, competition of this form is so ruthless that all consumer surplus is eliminated. Since workers know that once they become engaged in a competitive battle, all of their consumer surplus will be competed away, they will not sign contracts of this form (if there are contracts with positive consumer surplus available). When σ_ϵ^2 is small, the terms of the contest can be set in such a way that ruthless competition prevails only for some values of θ, while for other (smaller) values of θ, a conventional contest occurs. Contracts entailing some piece rate compensation can also eliminate such quandaries (see Gilbert and Stiglitz (1979) for a more extended discussion in the context of patent races).

References

AKERLOF, G. "The Economics of Caste and of the Rat Race and Other Woeful Tales." *Quarterly Journal of Economics,* Vol. 90 (1976), pp. 599–617.

——— AND SOSKICE, D. "The Economics of Sanctions." Mimeo, Department of Economics, University of California, Berkeley, December, 1976.

ARNOTT, R. AND STIGLITZ, J. "Equilibrium in Competitive Insurance Markets." Mimeo, Department of Economics, Princeton University, 1981.

ARROW, K. "Lecture Notes on Probability." Harvard University, 1970.

FARRELL, J. "A Sufficient Statistic Theorem." Mimeo, Department of Economics, M.I.T., 1980.

FITZROY, F. "Contests." Mimeo, International Institute of Management, West Berlin, 1981.

GILBERT, R. AND STIGLITZ, J. "Entry, Equilibrium, and Welfare." Mimeo, Department of Economics, Princeton University, 1979.

GREEN, J. AND STOKEY, N. "A Comparison of Tournaments and Contests." Discussion Paper 862, Harvard Institute of Economic Research, Harvard University, 1981.

GROSSMAN, S. AND HART, O. "An Analysis of the Principal-Agent Problem." Cambridge University Working Paper, 1980.

HART, O. "The Market Mechanism as an Incentive Scheme." Economic Theory Discussion Paper 53, University of Cambridge, 1981.

HOLMSTROM, B. "On Incentives and Control in Organizations." Ph.D. Dissertation, Stanford University, 1977.

———. "Moral Hazard in Teams." *Bell Journal of Economics,* Vol. 13, No. 2 (Autumn 1982), pp. 324–340.

LAZEAR, E. AND ROSEN, S. "Rank Order Tournaments as Optimum Labor Contracts." *Journal of Political Economy,* Vol. 89 (1981), pp. 841–864.

MIRRLEES, J. "Notes on Welfare Economics, Information, and Uncertainty" in M. S. Balch, D. L. McFadden, and S. Y. Wu, eds., *Essays on Economic Behavior under Uncertainty,* Amsterdam: North-Holland/American Elsevier, 1974.

———. "The Theory of Moral Hazard and Unobservable Behavior." Mimeo, Nuffield College, 1975.

MOOKHERJEE, D. "Optimal Incentive Contracts in Multiagent Situations." Ph.D. Dissertation, London School of Economics and Political Science, 1981.

NALEBUFF, B. "Prizes and Incentives." D. Phil. Thesis, Nuffield College, 1982.

——— AND STIGLITZ, J. E. "Information, Competition and Markets." *American Economic Review* (May 1983).

——— AND VARIAN, H. "Some Aspects of Risk Sharing in Nonclassical Environments." University of Michigan Working Paper, 1981.

O'KEEFE M., VISCUSI, K., AND ZECKHAUSER, R. "Economic Contests." Mimeo, Kennedy School of Government, Harvard University, 1982.

ROTHSCHILD, M. AND STIGLITZ, J. "Increasing Risk. I: A Definition." *Journal of Economic Theory,* Vol. 2 (1970), pp. 225–243.

STIGLITZ, J. "Incentives and Risk Sharing in Sharecropping." *Review of Economic Studies,* Vol. 41 (1974), pp. 219–256.

———. "Incentives, Risk, and Information: Notes towards a Theory of Hierarchy." *Bell Journal of Economics,* Vol. 6, No. 2 (Autumn 1975), pp. 552–579.

———. "Competition and Cooperation: Towards a General Theory of Compensation." Preliminary Work, Princeton University, 1980.

———. "Theory of Competition, Incentives, and Risks." Paper presented at the IEA Conference in New Developments in the Theory of Market Structures, Ottawa, May 1982, and forthcoming in the Proceedings, edited by J. E. Stiglitz and F. Mathewson.

——— AND WEISS, A. Part I, Part II: "Credit Rationing in Markets with Imperfect Information, Part I." *American Economic Review* (June 1981); Part II published as Princeton University Econometric Research Program Research Memorandum 268, August 1980.

VARIAN, H. "Redistributive Taxation as a Social Insurance." *Journal of Public Economics,* Vol. 14 (1980), pp. 49–68.