

American Economic Association

Productivity Under Group Incentives: An Experimental Study

Author(s): Haig R. Nalbantian and Andrew Schotter

Source: *The American Economic Review*, Vol. 87, No. 3 (Jun., 1997), pp. 314-341

Published by: American Economic Association

Stable URL: <http://www.jstor.org/stable/2951348>

Accessed: 22/06/2010 08:58

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/action/showPublisher?publisherCode=aea>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



American Economic Association is collaborating with JSTOR to digitize, preserve and extend access to *The American Economic Review*.

<http://www.jstor.org>

Productivity Under Group Incentives: An Experimental Study

By HAIG R. NALBANTIAN AND ANDREW SCHOTTER*

This paper presents an experimental examination of a variety of group incentive programs. We investigate simple revenue sharing and more sophisticated, target-based systems such as profit sharing or productivity gainsharing, as well as tournament-based and monitoring schemes. Our results can be characterized by three facts: (1) history matters; how a group performs in one incentive scheme depends on its history together under the scheme that preceded it; (2) relative performance schemes outperform target-based schemes; and (3) monitoring can elicit high effort from workers, but the probability of monitoring must be high and, therefore, costly. (JEL J33, C92)

Despite its recent resurgence, the productivity of American workers remains an issue of central concern of business and public policy. Traditionally, the efforts to strengthen American competitiveness have stressed technological advance and investment in physical and human capital. More recently, however, attention has been turning to the behavioral dimensions of labor productivity, the variations in the quantity and quality of labor inputs that stem from the complex of financial and non-financial inducements that constitute an organization's reward system. It is increasingly recognized in industry that by introducing carefully crafted group incentive compensation systems, it may be possible to induce American workers to work both harder and smarter and to use even existing technologies in new and better ways that enhance their productivity. In the short run at least, and perhaps even longer term, this may be the most effective instrument for raising productivity, yet insufficient attention has been paid to this possibility in the economics literature.

The paucity of empirical research in this area is quite surprising given the substantial theoretical advances that have been made in the analysis of labor contracting. The work in contract theory has shed considerable light on the nature of optimal contracts under alternative assumptions about the level and distribution of information amongst contracting parties and differences in their respective attitudes towards risk. But until very recently, there has been little empirical testing of the pertinent theoretical results in relation to group incentives. What has emerged in the last several years are a number of rigorous econometric studies of the relationship between profit sharing (and/or employee stock ownership plans) and labor productivity. (For an excellent and thorough review of this econometric literature, see Martin L. Weitzman and Douglas L. Kruse, 1990; also, U.S. Department of Labor, 1993.) But to our knowledge there has been no direct investigation of the *relative* performance of alternative *types* of group incentive systems, something which is of considerable academic and practical interest given the huge variety of group incentive systems actually employed by firms. As a result, there is little empirical basis for discriminating amongst the various group incentive programs, especially as concerns the structure of the sharing rules or payoff formulae that distinguish them.

Perhaps the reason for this lack of attention has to do with the difficulty of using naturally occurring data to answer these questions. Natural data on employee compensation and pro-

* Nalbantian: William M. Mercer, Incorporated, 1166 Avenue of the Americas, New York, NY 10036; Schotter: Department of Economics, New York University, 269 Mercer Street, New York, NY 10003. The authors would like to thank the C.V. Starr Center for Applied Economics for its financial and technical assistance in writing this paper. We also would like to thank Ken Rogoza for his help, as well as Blaine Snyder, Sunando Sen, and Tao Wang, who took a key role in helping us process the data. Finally, the comments of three anonymous referees are gratefully acknowledged.

ductivity are difficult to obtain and generally quite suspect in terms of quality. Moreover, the data required to undertake satisfactory empirical tests of the underlying theory include preference and production function parameters that are not directly observable. Beyond these factors there are the inherent limitations of statistical techniques with respect to endogeneity and "control" problems, limitations that are especially severe when comparisons are to be made of systems operating in diverse and highly idiosyncratic environments. These limitations are compounded in our case, as we are testing several mechanisms that are theoretically inspired and which exist, at best, infrequently (if at all) in actual practice. The experimental approach, on the other hand, readily allows for a systematic analysis of *ceteris paribus* changes in discrete aspects of given institutions, in our case, the different reward formulae. By so isolating the effects of these changes, we are better able to assess if the incentive properties identified in theory materialize in the actual behaviors of individuals operating in the stylized markets created in the lab.

In this paper we take a first, experimental step in rectifying this omission. We report on a set of nine experiments run using 408¹ paid human volunteers, the purpose of which was to investigate the problem of group moral hazard and the performance characteristics of several different classes of group incentive schemes now in use or deduced from economic theory. Unlike previous studies, we focus directly on the reward formulae themselves, trying to illuminate the behavioral and operational mechanics of the incentive structures they define. The specific schemes we investigate range from simple revenue sharing (egalitarian partnerships) to more sophisticated, target-based systems such as profit sharing and so-called productivity gainsharing.² These are prototypical real-world incarnations of what economists call "forcing contracts."

¹ We actually ran 13 experiments using 588 subjects, but a number of experiments were run to answer questions which subsequently were edited out of the paper.

² Productivity gainsharing is a form of group incentives in which employees share (usually 50–75 percent) in improvements over past performance. The programs typi-

We also examined the properties of team-based tournaments, denoted here as "competitive teams," in which intrafirm competition, for example, between profit centers, is created so that relative performance becomes the basis of incentive awards. The performance of all these group incentive systems then is compared to that of individual incentive systems characterized by probabilistic monitoring and efficiency wages.

The conclusions of our study are straightforward and can be summarized by four simple observations.

Observation 1: Shirking happens.—When experimental subjects are placed under an incentive plan which provides strong incentives to shirk, their effort levels do approach the shirking equilibrium as they near the end of the experiment.

Observation 2: History matters.—The performance of an experimental group using any particular group incentive formula depends on the effort norm established by this group in its previous experience with other incentive schemes. In addition, when a past common group experience is extremely positive (i.e., exhibits high levels of group output), current output levels tend to be high as well.

Observation 3: A little competition goes a long, long way.—Tournament-based group incentive mechanisms that create competition between subgroups in the organization for a fixed set of prizes (i.e., which create internal tournaments) determine higher mean outputs than all target-based mechanisms examined and smaller variances in those outputs than many of them.

Observation 4: Monitoring works but is costly.—When monitoring is possible but not perfect, high levels of effort can be elicited

cally are applied to production units and use measures of operating performance rather than broader financial measures as the basis of awards. Unlike profit sharing where targets are simply stipulated by management, under productivity gainsharing the historical performance of the group in some agreed-upon base period becomes the benchmark against which current performance is measured.

from workers. However, unless the probability of detection is great (and, therefore, costly to maintain), such monitoring schemes are likely to fail.

In this paper we proceed as follows. In Section I we motivate our work by presenting a quick overview of the industrial use of group incentive or variable pay schemes in the United States and the empirical research estimating their effectiveness. In Section II we make our discussion more precise by presenting the theory behind the exact group incentive mechanisms we are considering. Section III reviews our experimental design, while Section IV presents our results substantiating the four observations listed above. Section V offers some conclusions. Finally, an Appendix containing the instructions for our profit-sharing experiment is presented.

I. Industrial Practice and Previous Empirical Research

The use of group incentive or variable pay schemes has grown rapidly in the United States over the past 50 years. While in 1945 there were only 2,113 qualified deferred and combination-cash profit-sharing plans operating in the United States, by 1991 this figure had risen to 490,000, with over one-quarter of them including immediate cash payments. Carla O'Dell and Jerry L. McAdams (1987) found that 13 percent of firms responding to their survey has some form of gainsharing in place. With respect to employee stock ownership programs (ESOPs), Joseph R. Blasi and Douglas L. Kruse (1991) project that by the year 2000 more than one-quarter of publicly traded firms on the New York, American, and over-the-counter Stock Exchanges will be more than 15 percent owned by their employees.

The bulk of empirical research on group incentives is represented either by field studies detailing company experiences with specific group incentive plans (for example, see National Commission on Productivity and Work Quality, 1975; Mitchell Fein, 1982; Brian E. Graham-Moore and Timothy L. Ross, 1983) or by simple correlational studies examining the relationship between the adoption of group incentive plans and various measures of firm performance and/or labor productivity (see

Bion B. Howard and Peter O. Dietz, 1969; Betram L. Metzger and Jerome A. Colletti, 1971; Metzger, 1975). In addition, a number of industrial surveys have been conducted that compare and evaluate the effectiveness of broad classes of group incentive plans (U.S. General Accounting Office, 1981; New York Stock Exchange Office of Economic Research, 1982; McAdams and Elizabeth J. Hawk, 1994). Overwhelmingly, the assessments of group incentives offered in these studies are positive, though the variance of productivity effects is considerable.

Some of the more recent studies have utilized more sophisticated statistical techniques applied to time-series and cross-sectional data (for example, see Felix FitzRoy and Kornelius Kraft, 1986, 1987; John A. Wagner III et al., 1988; John Cable and Nicholas Wilson, 1989; Kruse, 1992). These studies show strong positive effects of group incentives on various productivity and financial performance variables even, in some cases, after addressing the problems of endogeneity. Group incentives also have been considered in several recent empirical studies of the effects on productivity of alternative human resource "systems" (for example, see Mark Huselid, 1995; Casey Ichniowski and Kathryn Shaw, 1995; Ichniowski et al., 1996). These studies show that the more participatory work systems emphasizing decentralized decision-making, extensive information sharing, flexible job assignments, and some form of profit sharing, among other things, tend to outperform traditional hierarchical system designs. Of course, there are exceptions to these findings (Jone L. Pearce et al., 1985; Andrew Weiss, 1987; Michael Conte and Jan Svejnar, 1989), and much of the empirical literature cited is suspect due to methodological weaknesses, most notably, the frequent failure to control for other potential explanatory variables and for feedback relationships. Nonetheless, it is clear that the preponderance of evidence supports the claim that group incentives can, and often do, contribute to significant increases in labor productivity and firm performance.³

³ Detailed reviews and evaluations of the empirical evidence can be found in Nalbantian (1987) and Alan S. Blinder (1990).

To our knowledge, except for the work of Schotter and Keith Weigelt (1992a) on long- and short-term corporate bonuses, and Clive Bull et al. (1987) and Schotter and Weigelt (1992b) on symmetric and asymmetric tournaments, there has not been any experimental work in economics on the group incentive problem as posed herein. However, the problem posed here shares many common characteristics with the public goods problem, and to that extent a rich body of pertinent experimental work does exist. We will comment on the links between this work and ours in Section II below.

II. A Simple Model of the Group Incentive Problem

To economize on space, we will present a set of models which underlie our experiments, using the exact experimental parameterization and functional forms. This will allow us to avoid presenting the theory twice—once in its general, and once in its specific, experimental form. Consider a firm composed of six workers indexed $i = 1, 2, \dots, 6$. Each member of the firm can choose an effort level e_i from the closed interval $[0, 100]$. Effort is costly, with the cost defined by $C(e_i) = e_i^2/100$. The effort levels of the firm's workers produce output using a simple stochastic linear technology specified as:

$$(1) \quad Y = \sum_1^6 e_i + \varepsilon,$$

where Y is firm output, e_i is the effort of the i th worker, and ε is a random variable defined uniformly over the integers in the closed interval $[-40, +40]$. Assume next that this firm sells its output on a competitive market for a price of 1.5. As a result, the firm's revenue is

$$(2) \quad R = 1.5Y = 1.5 \left(\sum_1^6 e_i + \varepsilon \right).$$

Given this specification, the Pareto-optimal effort level for each worker can be defined by solving the following simple maximization problem:

$$(3) \quad \max \pi = 1.5 \left(\sum_1^6 e_i + \varepsilon \right) - \sum_1^6 e_i^2/100.$$

Given R is linear in each e_i and each $C(e_i)$ is strictly convex, the first-order conditions define a unique profit-maximizing effort level as:

$$(4) \quad \partial \pi_i / \partial e_i = 1.5 - 2e_i/100 = 0, \\ i = 1, 2, \dots, 6, \quad \text{or} \quad e_i = 75.$$

The problem for principal-agent theory is how to design an incentive scheme or mechanism which will implement these Pareto-optimal effort levels as Nash equilibria.

In the remainder of this section we will review a set of different incentive mechanisms which can either be derived from principal-agent theory or actually observed in industrial practice. Some of these mechanisms implement Pareto-optimal effort levels as Nash equilibria while others do not. In general, we lump all incentive schemes into four categories: partnership schemes, target-based schemes, tournament-based schemes, and individualistic monitoring schemes. We will review these types of mechanisms each in turn since they are the mechanisms we eventually test experimentally.

A. Partnership Schemes: Revenue Sharing

Egalitarian partnership schemes, comparable to the voluntary contribution mechanisms of public goods theory, represent the archetypical incentive mechanism for which free-riding is a dominant strategy (see Section II, subsection E, for a comparison of public goods and group incentive experiments). As such, they suffer from the same disincentive effects. We illustrate such schemes using a revenue-sharing mechanism in which the payoff to the i th worker is defined as follows:

$$(5) \quad \pi_i = 1.5 \left(\sum_1^6 e_i + \varepsilon \right) / 6 - e_i^2/100.$$

As we can see, in this scheme all revenue generated by the firm is shared equally and a worker's final payoff is simply his revenue share minus his cost of effort.

To determine the Nash equilibrium effort levels defined by the game associated with this mechanism, we take the partial derivative of

each worker's payoff function with respect to his or her own effort level e_i and set it equal to zero. This yields:

$$(6) \quad \partial\pi_i/\partial e_i = 1.5/6 - 2e_i/100 = 0$$

$$\text{or} \quad e_i = 12.5.$$

Note that since the marginal benefit from exerting a unit's worth of effort is 1.5/6 and is independent of the effort level, choosing 12.5 is a dominant strategy under egalitarian revenue sharing. Hence, this scheme defines a typical free-riding prisoner's dilemma situation in which there is a dominant strategy yielding Pareto-inferior outcomes for all.

B. Target Based Schemes Forcing Contracts

Bengt R. Holmstrom (1982) has made a number of suggestions about finding solutions to the shirking dilemma presented by revenue sharing. Among his suggestions is the forcing contract mechanism. Such a mechanism is the generic form representative of a class of target-based mechanisms in which revenue or other outcome targets are set exogenously for the firm or a performance group within the firm. If the target revenue is achieved, the workers share in all of the revenue generated, while if the firm's revenue falls short of the target, each worker is paid a relatively low penalty wage.

More formally, the payoff to workers under this kind of forcing contract is defined as follows:

$$(7) \quad \pi_i = \begin{cases} 1.5 \left(\sum_1^6 e_i + \varepsilon \right) / 6 - e_i^2 / 100 & \text{if } 1.5(\sum e_i + \varepsilon) \geq R^* \\ B & \text{otherwise.} \end{cases}$$

Such forcing contracts have many Nash equilibria, each characterized by a different $Y^* - B$ pair. To find these Nash equilibria, let $P(e_i, \sum_{j \neq i} e_{-i})$ denote the probability that a group meets its target of Y^* given an effort level of e_i by agent i and $\sum_{j \neq i} e_{-i}$ by the other agents excluding i . Note that for a fixed e_i and $\sum_{j \neq i}$

e_{-i} , the expected value of the firm's output, conditional on meeting the target, is

$$(8) \quad E(Y | Y > Y^*) = \frac{\left(e_i + \sum_{j \neq i} e_j + Y^* + 40 \right)}{2},$$

where the constant 40 represents half of the support of the random variable ε .

Consequently, each worker faces a payoff function of

$$(9) \quad \pi_i \left(e_i, \sum_{j \neq i} e_j \right) = B + P \left(e_i, \sum_{j \neq i} e_j \right) \times \left[\frac{1.5 \left(e_i + \sum_{j \neq i} e_j + Y^* + 40 \right)}{2} - B \right] - \frac{e_i^2}{100}.$$

For a Nash equilibrium the following first-order condition must hold for each i :

$$(10) \quad \frac{\partial\pi_i}{\partial e_i} = -P'(\cdot)B + P'(\cdot) \left[\frac{1.5 \left(e_i + \sum_{j \neq i} e_j + 40 + Y^* \right)}{2} \right] + P(\cdot)(0.125) - \frac{2e_i}{100} = 0,$$

$$i = 1, 2, \dots, 6.$$

This condition sets up a relationship, given all of the other parameters, between Y^* and B such that in order to implement $Y^* = 450$ ($e_i^* = 75$) as a Nash equilibrium, we must set $B \leq 1.125$.

Profit Sharing

It should be obvious to the reader that what we happen to call a profit-sharing scheme is nothing more than a forcing contract scheme with a lower target and a penalty wage of zero. Our original intention was to devise a profit-sharing scheme characterized by a base wage and target consistent with the shirking equilibrium of the revenue-sharing experiment. In short, our aim was to set a target $Y^* = 75$ ($R^* = 112.5$) and a base wage of $112.5/6 = 18.75$. This would guarantee subjects a wage equal to their expected equilibrium payout of the revenue-sharing scheme and then allow them to share revenue as it increased above the target of 112.5. If output actually rose above this level, the increase could be attributable to the profit-sharing aspect of the scheme. Unfortunately, by guaranteeing subjects a wage of 18.75 and setting the target at $Y^* = 75$, we also guaranteed a dominant strategy for subjects to shirk and exert zero effort. Hence, we could not create a prototypical profit-sharing scheme enforcing the revenue-sharing Nash equilibrium without simultaneously altering some other institutional variable as well (e.g., introducing an individualistic incentive component like monitoring). To maintain the condition of allowing discrete, *ceteris paribus* changes only, we instead devised what we consider to be a second-best experiment which kept the target at the revenue-sharing equilibrium $Y^* = 75$ ($e_i^* = 12.5$) but lowered the penalty wage B to zero ($B = 0$). With these parameters, the Nash equilibrium involves each subject choosing 19.1. For these reasons we call this experiment a profit-sharing experiment—yet that designation is arbitrary.

Gainsharing

A gainsharing scheme is a target-based, profit-sharing scheme in which the target is generated endogenously by the previous output of workers. Hence, gainsharing is a forcing contract with a target based on historical or “base period” performance. In our experiments we always perform gainsharing after an initial 25-round revenue-sharing experiment, taking the average output of the last ten rounds of the revenue-sharing experiment as the target for the gainsharing experiment. (Subjects were not told that their

performance in the first revenue-sharing experiment would in any way influence their payoffs in the second experiment they would perform. In fact, they knew no details at all of the second gainsharing experiment they were to perform until the revenue-sharing experiment that preceded it was completed, though they knew that a second experiment would occur.) Note that since we do not know, *a priori*, the output of our subjects in their revenue-sharing experiment, we could make no predictions about the output for these gainsharing experiments, or whether they even had equilibria.

C. Tournament-Based Schemes: Competitive Teams

In contrast to target-based schemes, tournaments make the payoffs of agents or groups of agents contingent upon relative, rather than absolute, performance. In our experiments we test a tournament-like mechanism which we call competitive teams, which involves dividing the firm into two (or more) teams and having these teams compete for prizes. The team producing the most output gets the big prize, while the loser gets the small prize. As a result, our competitive teams scheme relies on competition to motivate the workers functioning under it.

To be more precise, let the firm be divided into two teams, T_1 and T_2 , and let $R_1 = 1.5(Y_1)$ and $R_2 = 1.5(Y_2)$ be the revenues and outputs generated by these teams. Under a competitive team mechanism, the payoff for any worker i on Team 1 is defined as:

$$(11) \quad \pi_i(Y_1, Y_2, e_i)$$

$$= \begin{cases} \frac{R_1 + TR}{6} - \frac{e_i^2}{100} & \text{if } Y_1 > Y_2 \\ \frac{R_1 - TR}{6} - \frac{e_i^2}{100} & \text{if } Y_1 < Y_2, \end{cases}$$

where TR is a transfer made from the winning team to the losing team. A similar payoff function can be defined for workers on Team 2.

Note that as formulated above and as implemented in the lab, the competitive team scheme is played as a noncooperative game with each worker choosing its effort level in

isolation and without knowledge of the choice made by its team members. In the experiments only the teams' revenues, and not any individual member's effort, were announced to the subjects after each round of the experiment. Also, note that we specified the mechanism as defining a transfer made from the winning team to the losing team. Clearly, we easily could have specified the winning team as receiving a bonus paid by the firm and not as an intrafirm transfer from one branch to another. Economically, but perhaps not psychologically, they are equivalent.

Let $E_1 = \sum_{i \in T_1} e_i$ for all workers on Team 1 and $E_2 = \sum_{i \in T_2} e_i$ for all workers on Team 2.

To demonstrate that the competitive team mechanism is capable of implementing Pareto-optimal outcomes as Nash equilibria, assume that all members of each team choose an effort level of 75. Given these choices, the expected revenue for each team is 675 and the probability of each team winning the transfer is $1/2$. Now consider one worker, say on Team 1, who contemplates a change in his or her effort by a marginal amount. Increasing effort marginally increases the probability of winning by $\partial Pr/\partial e_i$. The benefit of winning is $(2 \times TR)/6 + (\partial R/\partial e_i)/6$, i.e., the difference between winning the transfer and losing it, $2TR$, and one's share in the marginal revenue generated for the team $(\partial R/\partial e_i)/6$. The marginal cost of changing one's effort is $2e_i/100$. Hence, in order for each worker not to want to deviate from the Nash equilibrium we must have

$$\begin{aligned} \partial Pr(\cdot)/\partial e_i \cdot [(2 \times TR)/6] \\ + (\partial R/\partial e_i)/6 = 2e_i/100. \end{aligned}$$

Given our experimental parameters ($\varepsilon \in [-40, +40]$, $p = 1.5$, and $C(e_i) = e_i^2/100$), if TR is set equal to 360, a Nash equilibrium exists in which all subjects choose 75.

D. Individualistic Schemes: Monitoring

Our final incentive scheme is not a group incentive scheme at all, but an individual wage-cum-supervision mechanism (see Guillermo A. Calvo and Stanislaw H. Wellisz, 1978; Calvo, 1987). Under this mechanism our firm offers its workers a wage W greater

than their opportunity wage w if they put out an effort of e^* when on the job. The firm will check the effort level of the worker with a probability of p each period; if the worker is caught working at an effort level lower than e^* , he or she will be fired. Again, effort is assumed to be costly for the worker as defined by the cost function $C(e_i) = e_i^2/100$. In short, the worker is offered an efficiency wage if he or she will put out an effort level of e^* .

With correctly set parameters, this scheme will determine Pareto-optimal effort levels on the part of the workers. Obviously, a worker will shirk depending on whether the expected payoff from shirking is greater than that of working. Realizing that the optimal shirk sets $e_i = 0$, while the optimal nonshirking effort level sets $e_i = e^*$, we see that shirking will occur if

$$\begin{aligned} (12) \quad E\pi_i(e_i = 0) \\ = p \times (w) + (1 - p) \times W \geq W \\ - (e^*2)/100 = E\pi_i(e_i = e^*) \\ \text{or} \quad (W - w) \geq [(e_i^2)/100] \times \frac{1}{p}. \end{aligned}$$

Setting $p = 0.70$, $W = 112.5$, $w = 18.75$, and $e^* = 75$, we see that the worker would prefer to work rather than shirk. When $p = 0.30$, the opposite conclusion arises. There is a direct relation between the intensity of supervision and the size of the efficiency wage required to deter shirking.

To summarize, we present Table 1, which also furnishes us predictions for the experiments to be reported on in Section IV. Note that for all schemes, the Pareto-optimal effort level is 75 for each worker, which yields an individual payoff of 56.25, a group output of 450, a group revenue of 675, and a group profit (group revenues minus group effort costs) of 337.5.

E. Relationship to the Public Goods Literature

There is an obvious analogy between public goods experiments and our group incentive experiments; in the latter, group output is, in

TABLE 1—THEORETICAL PREDICTIONS*

Experiment	Effort	Payoff	Group output (revenue)	Profit (revenue— group effort cost)
1. Revenue sharing	12.5	17.18	75 (112.5)	103.1
2. Forcing contract (75)	75	3.5	450 (675)	337.5
3. Forcing contract (40)	40	19.41	240 (360)	264
4. Profit sharing $B = 0$ $Y^* = 112.5$	19.1	25	114.6 (171.9)	150.01
5. Gainsharing $B = 0$ $Y^* = ?$	**	**	**	**
6. Competitive teams Transfer = 360	75	56.2	450 (675)	337.5
7. Monitoring $p = 0.70$ $W = 112.5$ $w = 18.75$	75	56.25	450 (675)	337.5
8. Monitoring $p = 0.30$ $W = 112.5$ $w = 18.75$	0	84.32	0 (0)	-506.25

*Note that for all mechanisms there are six subjects facing a cost of effort function of the form $c(e_i) = (e_i)^2/100$. Production is of the form $x = \Sigma e_i + \varepsilon$. ε will be distributed uniformly over the interval $[-40, +40]$, and all subjects will choose their effort from the closed interval $[0, 100]$. Finally, when subjects divide team revenue they do so equally so that each worker gets a share of 1/6th.

**Since we do not know what Y^* will be in the gainsharing experiment until we have the historical data from which to calculate it, we cannot know in advance what the Nash equilibrium of the gainsharing experiment is. This fact is indicated by the asterisks in the table.

fact, a kind of public good which is nonexcludable and shared equally by the workers. Despite this similarity, some differences do exist, since the environment defining some group incentive experiments is different from those typically seen in public goods experiments, as are some of the mechanisms or schemes we investigate. Let us review these differences each in turn.

With respect to environments, while group output is a public good, it is exhaustible and subject to crowding, since as n grows any fixed amount of group output gets shared by more and more people. Further, our group output is stochastic in the sense that for any level of group effort the eventual output (and, there-

fore, the probability of surpassing a target) is stochastically determined by the addition of the random shock ε . Both of these elements are missing in most public goods experiments, yet could easily be incorporated. Finally, while most public goods environments specify a demand function for the public good for each subject (or at least a marginal willingness to pay function) with a constant marginal cost, in our experiments subjects face a convex and increasing cost function with a constant marginal revenue of 1.5 for the group output.

With respect to mechanisms, although some of the group incentive mechanisms we investigate have natural interpretations as public goods mechanisms, others could not easily be

used in public goods contexts. For example, while revenue sharing can be interpreted as a straightforward voluntary contribution mechanism similar to that of R. Mark Isaac and James M. Walker (1988), and forcing contract mechanisms can be considered simple threshold mechanisms similar to those of Gerald Marwell and Ruth E. Ames (1980) and Thomas Palfrey and Howard Rosenthal (1991), mechanisms such as competitive teams and gainsharing have no easy analogue in public goods theory. More precisely, the technology of public goods construction does not readily permit communities to be split in two and compete in their contributions to a public good. The same is true of gainsharing, which is an inherently dynamic mechanism with thresholds or provision points set endogenously. Finally, while public goods threshold mechanisms are capable of offering money-back guarantees (see Vernon L. Smith, 1977; Robyn M. Dawes et al., 1986; Jeffrey S. Banks et al., 1988), such that if the public good is not built, all citizens get their money back and nothing is lost; in group incentive problems such guarantees are not natural since effort is expended and lost forever, whether or not the target is reached.

In summary, while group incentive experiments share many of the properties of public goods experiments, there are more than enough differences to call for them to be studied in their own right.

III. The Experiments and Experimental Design

A. *The Experiment*

To investigate our various group incentive formulae, we ran a set of nine different experiments using 408 college undergraduates recruited in groups of 12 from undergraduate economics courses at New York University. Students were requested to come to an experimental computer laboratory at the C.V. Starr Center for Applied Economics.⁴ They were paid \$3.00 for showing up and engaged in an experiment lasting about 1 hour and 20 minutes. Their average final payoffs for this

amount of time was about \$14.00–\$15.00, which seemed more than sufficient to motivate them. Because of the convex cost function, the marginal incentives to work or shirk are not linear and depend on effort levels. Furthermore, since different mechanisms implement different group output levels as equilibria, the marginal incentives vary across these equilibria as well. However, we know from the definition of Pareto-optimality that any of our mechanisms which implement Pareto-optimal Nash equilibria provide identical marginal incentives for our subjects at those equilibria. For the revenue-sharing equilibrium, it is a dominant strategy to choose 12.5 since the marginal private revenue generated from exerting one more unit of effort is $1.5/6 = 0.25$ everywhere while, with increasing convex costs, the marginal cost of doing so is less than 0.25 for all $e_i \leq 12.5$ and greater than 0.25 for all $e_i \geq 12.5$ no matter what group output is.

The experiments engaged in were a direct implementation of the incentive plans described in Section II. For instance, in all experiments (except the monitoring experiments) the 12 subjects recruited were randomly divided into two different groups of six subjects who remained anonymously grouped during the entire experiment. After reading the experiment's instructions (and having them read aloud by an experimental administrator, who answered any questions), subjects were asked to type a number between 0 and 100 into their computer terminals. Such a number can be interpreted as their unobservable effort levels, although in the instructions only neutral language was used. After these numbers were entered by each subject, the program guiding the experiment added up all of these numbers for each group separately and drew a random number uniformly distributed between -40 and $+40$ independently for each group. The random numbers for each group were added to the sum of their effort levels. In the instructions subjects were told that the higher the decision number they chose the higher their costs would be, and they were given a cost table illustrating the cost of each integer between 0 and 100. (This table was an integer representation of the cost function $e_i^2/100$).

The payoffs for each experiment then were determined according to the incentive plan be-

⁴ The instructions for the experiments are contained in the Appendix of the paper.

ing used as discussed in Section II. After each round, subjects could see only their own effort levels and the output levels of their own group. (In the competitive team experiment, each team also could see the output of its opposing team after each round.) No information about the individual effort levels of other subjects ever was revealed.

When round 1 of the experiment was over, round 2 started and was identical to round 1. Each experiment lasted for 25 rounds, which we felt was a sufficient length of time to foster learning if any was to occur. After the first 25 rounds were over, new instructions were handed out for a second experiment. (Subjects were not told about the details of the second experiment before they engaged in the first, but were informed that some second experiment would occur. This was done so that no interexperiment strategies could be engaged in while informing subjects that they would be held in the laboratory for another experiment.) In the second experiment, all subjects stayed with their same group. The payoffs at the end of the experiment were simply the sum of the payoffs of the subjects over the 50 rounds of their experience. Payoffs in each round were made in terms of points, which were converted into dollars at a rate that was known in advance by all subjects.

B. *Experimental Design—Phase I and Phase II*

Since our original intent in running these experiments was to investigate the incremental impact of group incentive formulae on poorly functioning work units, Phase I of our experimental design involved running each group first in a revenue-sharing experiment with the expectation that we would observe the low effort equilibrium in which each subject chose an effort of 12.5. Using this as a baseline and a basis for comparison, we then had each group perform a second experiment where one of our four other group incentive plans (profit sharing, gainsharing, forcing contracts, and competitive teams) was used. Hence, in Phase I we ran a set of experiments, all of which started with 25 rounds of revenue sharing followed by 25 rounds of some other group incentive experiment. Because we also were

interested in observing how history affected group performance, in Phase II of our experiment we reversed the process and ran other groups of subjects for 25 rounds of some non-revenue-sharing experiment first, followed by 25 rounds of a revenue-sharing experiment. For monitoring experiments in Phase I we ran 25 rounds of a monitoring experiment with a detection probability of 0.70, followed by 25 rounds of the same experiment, but with a monitoring probability of 0.30. In Phase-II experiments the order of these experiments was reversed. Our design is presented in Table 2 below.

Note that in our experimental design all cost and production parameters are held constant so that in moving from one experiment to another only the incentive *formula* changes. This allows us to attribute all changes in behavior to the formula and not to differing cost or production parameter levels. We compare purely institutional changes.

IV. Results

As previously stated, we will present the results of our experiments by substantiating Observations 1–4.

A. *Observation 1: Shirking Happens*

To illustrate this observation we concentrate on the behavior of our subjects in the revenue sharing-experiments. We consider these experiments to be classic examples of pure shirking or free-riding experiments since they present subjects with a dominant strategy of choosing an effort level 12.5 which leads to Pareto-inferior outcomes. These experiments furnish us with a baseline from which to measure the effectiveness of the other plans; hence it is crucial to establish behavior here first.

Figures 1 and 2 present the mean and median effort levels of our revenue-sharing experiments when they were performed first (Phase I). As we can see, while both the mean and median effort levels of subjects in round 1 start off at 34.86 and 34.08, respectively, by round 25 they converge toward the equilibrium effort level of 12.5. While they remain above this level (18.63 and 17.67 for the mean and median, respectively, in round 25), there is a clear downward tendency in the data. This

TABLE 2—EXPERIMENTAL DESIGN
PHASE I

Experiment number	Contract type rounds 1–25	Contract type rounds 26–50	Number of groups	Number of subjects
1	Revenue sharing	Forcing contract	7	42
2	Revenue sharing	Competitive teams	10	60
3	Revenue sharing	Profit sharing	6	36
4	Revenue sharing	Gainsharing	10	60
5	Monitoring $p = 0.70$	Monitoring $p = 0.30$	2	15
PHASE II				
Experiment number	Contract type rounds 1–25	Contract type rounds 26–50	Number of groups	Number of subjects
6	Forcing contracts	Revenue sharing	10	60
7	Competitive teams	Revenue sharing	10	60
8	Profit sharing	Revenue sharing	10	60
9	Monitoring $p = 0.30$	Monitoring $p = 0.70$	2	15

data is consistent with the stylized facts about public goods experiments as described by John O. Ledyard (1995); he notes that a typical public goods experiment starts out with approximately a 50-percent (of Pareto-optimal levels) contribution rate and then decreases to approximately 11 percent as the experiment is repeated.

We take these results as supporting the hypothesis (as summarized in Observation 1) that subjects will take advantage of shirking opportunities when they function under a plan that at least poses these opportunities in the form of dominant strategies.

B. Observation 2: History Matters

To illustrate this observation, we first compare the mean and median effort levels of groups engaged in our profit-sharing, forcing contract, and competitive team experiments in Phases I and II of our experimental design (i.e., before and after subjects have a history with revenue-sharing schemes). Note that two of these mechanisms (forcing contracts and competitive teams) implement Pareto-optimal effort levels as Nash equilibria, while the other, profit sharing, entails a suboptimal Nash

equilibrium. Hence, we will be comparing the effort levels of subjects in Experiments 3 and 10, 1 and 7, and finally 2 and 9, and look to see if these before-and-after experiences differ. Figures 3–8 illustrate our conclusions here.

As we can see, for both the median and the mean, effort levels in practically all rounds were higher in those experiments run before revenue sharing. In other words, the experiences of groups in the revenue-sharing experiment (where effort levels tended to move toward the low effort Nash equilibrium level of 12.5) tended to lower the effort levels of groups in their second non-revenue-sharing experiment below what they were when those same schemes were run first in Phase II. Consequently, previous history with a revenue-sharing mechanism that encourages shirking leads to lower outputs with subsequent non-revenue-sharing mechanisms.

To investigate the impact of history more closely, we ran a set of dummy variable regressions. In the first regression (Regression 1), run on our Phase-I data, we attempt to explain the mean effort level of groups in the first five rounds of the second non-revenue-sharing experiment as a function of their history (mean

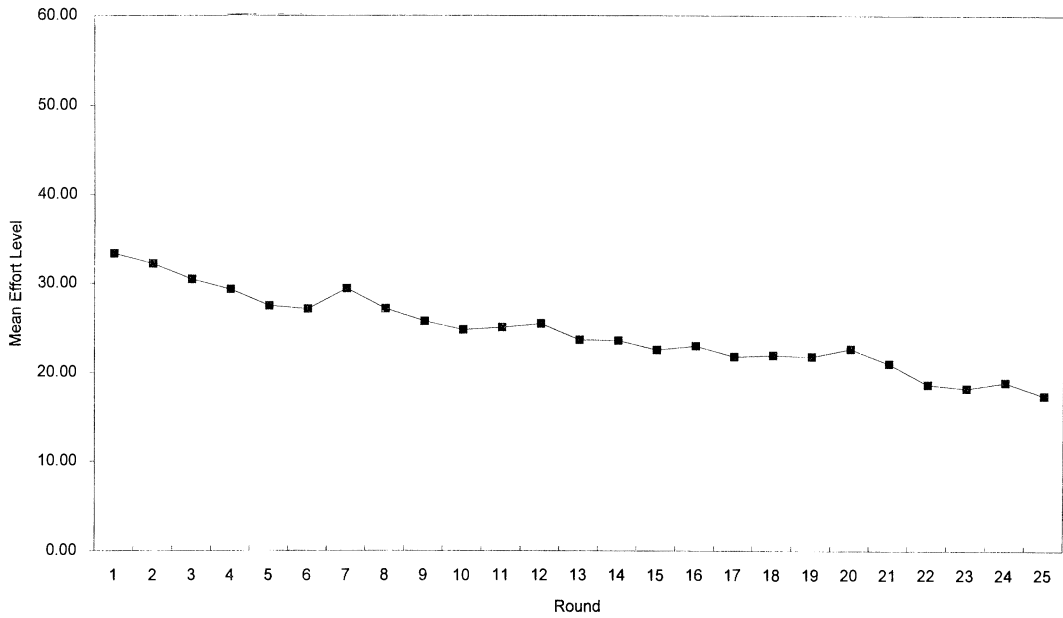


FIGURE 1. REVENUE SHARING (PHASE I)—MEAN EFFORT LEVELS

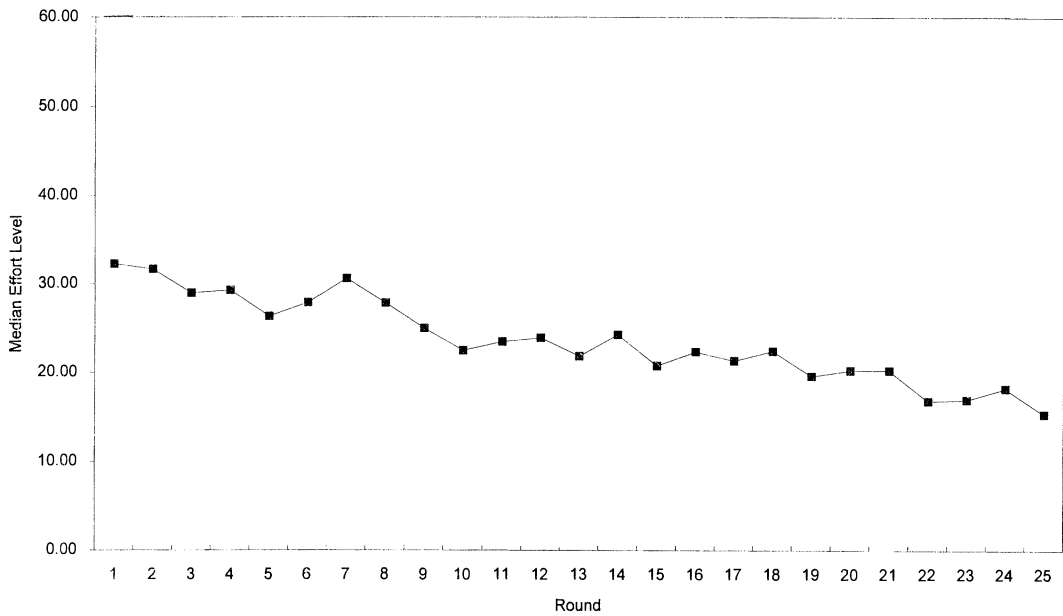


FIGURE 2. REVENUE SHARING (PHASE I)—MEDIAN EFFORT LEVELS

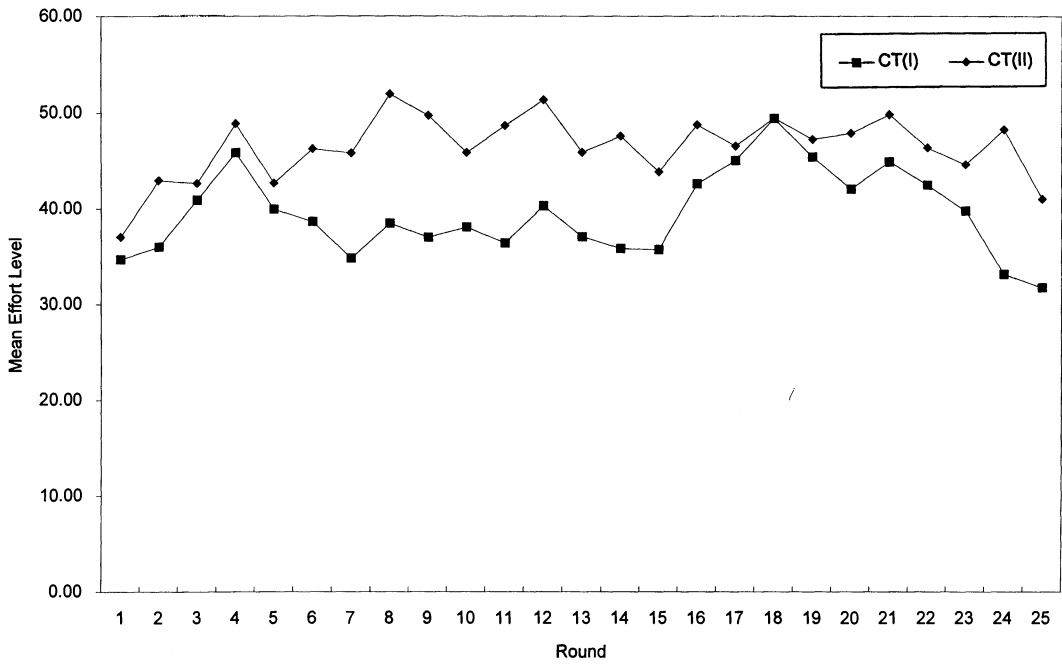


FIGURE 3. COMPETITIVE TEAMS—MEAN EFFORT LEVELS

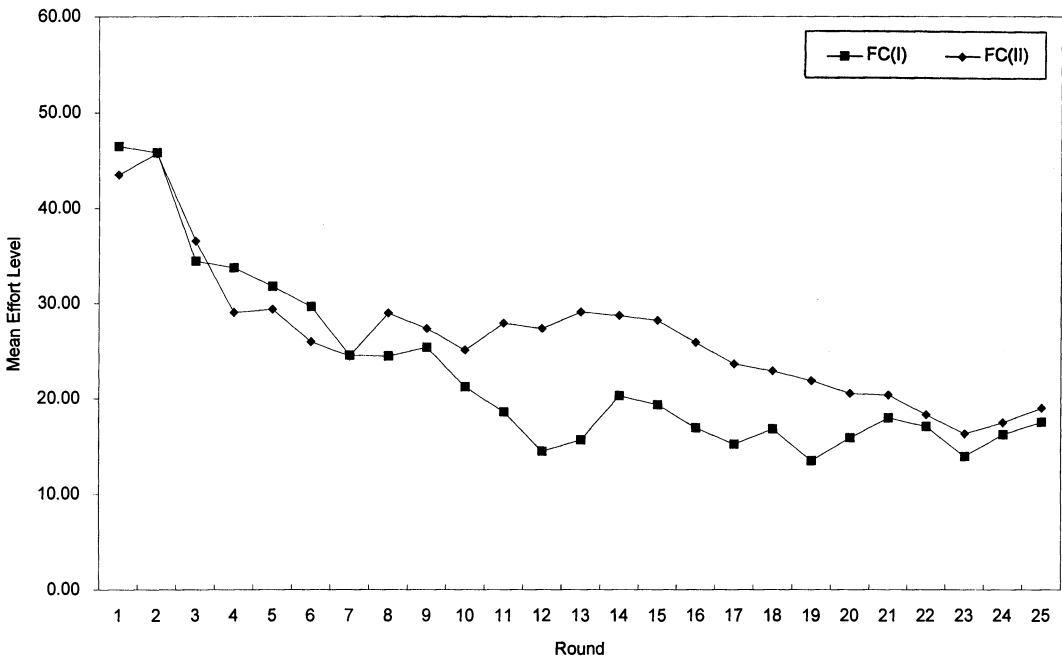


FIGURE 4. FORCING CONTRACTS—MEAN EFFORT LEVELS

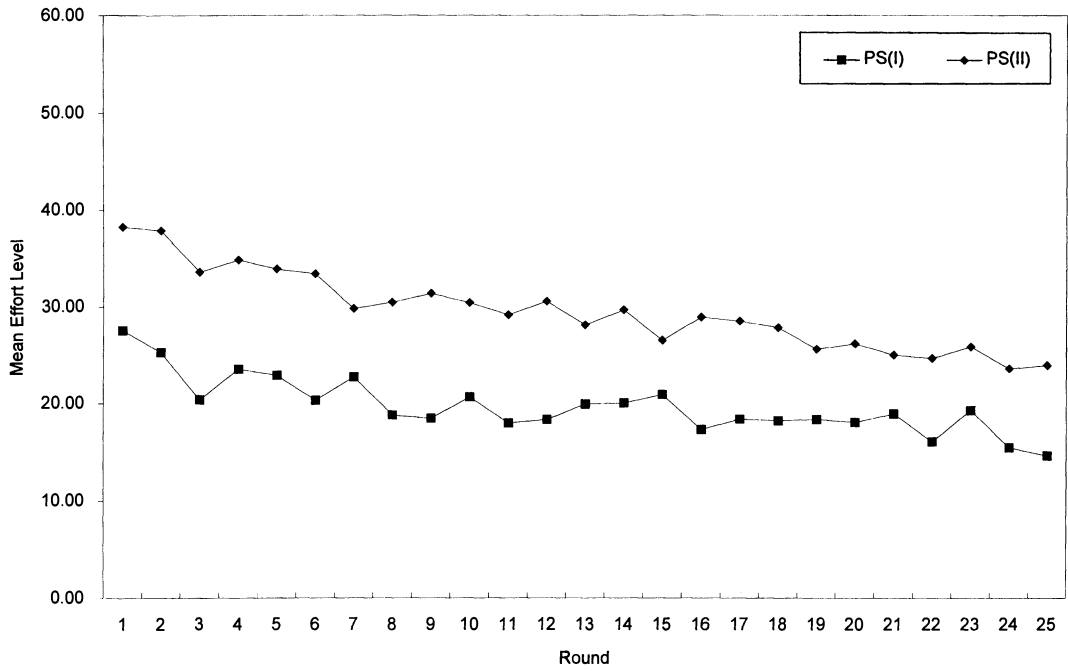


FIGURE 5. PROFIT SHARING—MEAN EFFORT LEVELS

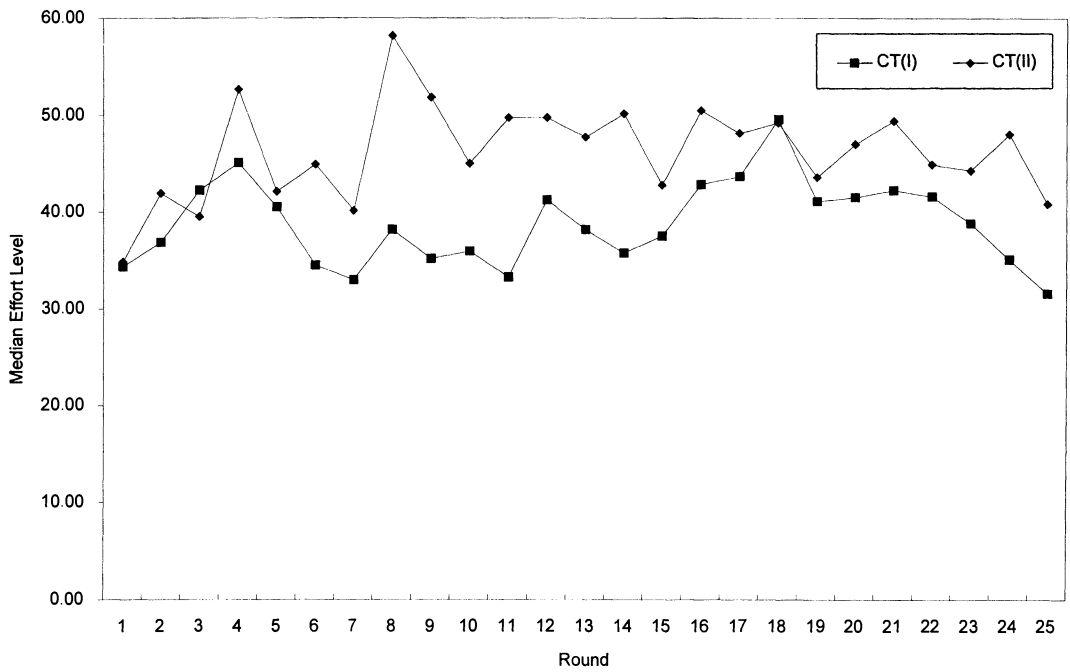


FIGURE 6. COMPETITIVE TEAMS—MEDIAN EFFORT LEVELS

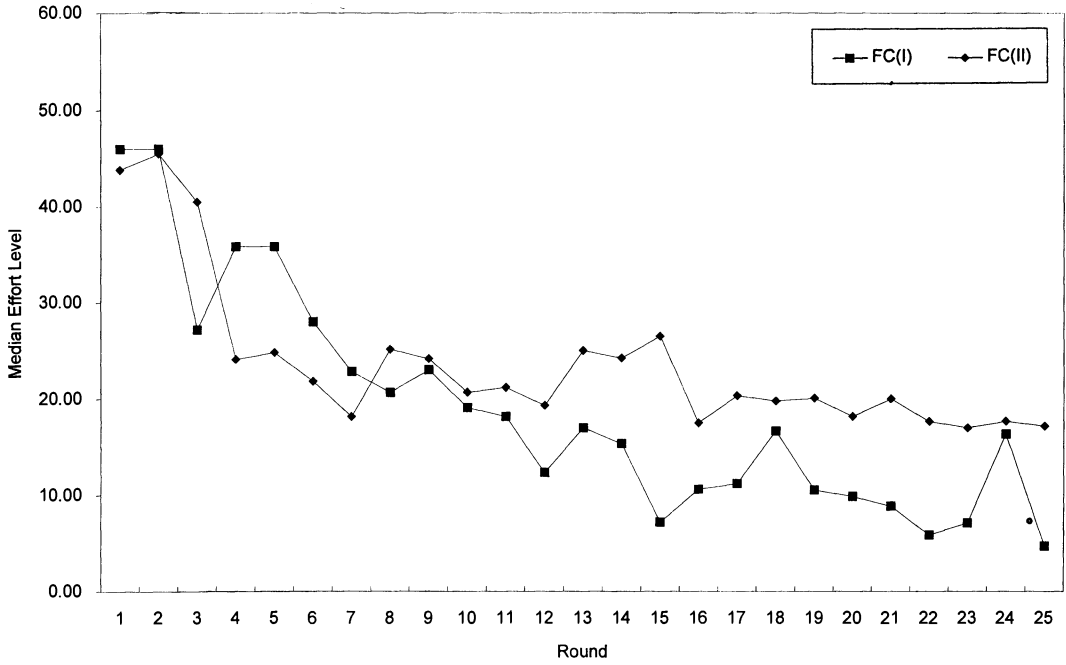


FIGURE 7. FORCING CONTRACTS—MEDIAN EFFORT LEVELS

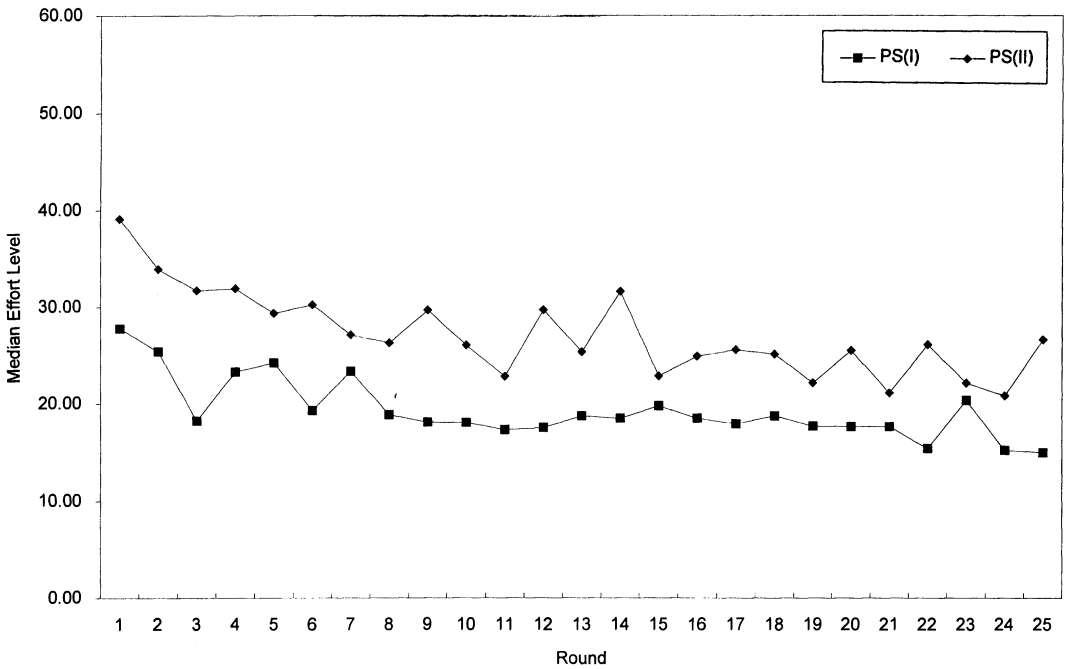


FIGURE 8. PROFIT SHARING—MEDIAN EFFORT LEVELS

TABLE 3—DUMMY VARIABLE REGRESSION

Variable	Regression 1	Regression 2	Regression 3
α (constant)	✓	✓	✓
D_1	✓	✓	✓
D_2	✓	✓	✓
D_3	✓	✓	✓
D_4	Dropped	✓	Dropped
D_5	N.A.	✓	N.A.
D_6	N.A.	✓	N.A.
D_7	N.A.	Dropped	N.A.
D_8	✓	✓	✓
D_9	✓	✓	✓
D_{10}	✓	Dropped	✓
D_{11}	Dropped	N.A.	Dropped
D_{12}	N.A.	N.A.	N.A.
D_{13}	N.A.	N.A.	✓
D_{14}	N.A.	N.A.	✓
D_{15}	N.A.	N.A.	✓
D_{16}	N.A.	N.A.	✓
D_{18}	N.A.	N.A.	✓
D_{19}	N.A.	N.A.	Dropped
D_{20}	N.A.	N.A.	N.A.
D_{21}	N.A.	N.A.	N.A.

individual effort level) in the last five rounds of their first revenue-sharing experiment and the current plan being used. A second dummy variable regression (Regression 2) was run using Phase-II data in an attempt to explain the mean effort level of groups in the first five rounds of the second revenue-sharing experiment as a function of their history (mean individual effort level) in the last five rounds of the first non-revenue-sharing experiment and the plan used previously. Finally, a third regression (Regression 3) was run seeking to explain the mean effort level of groups in the last five rounds of the second non-revenue-sharing experiment as a function of the plan used, the mean effort levels of the group during the last five rounds of the previous experiment, and the mean effort levels of the first five rounds of the current one.

The following dummy variables were defined:

D_i is a dummy variable taking a value of 1 if the experience of the group (its mean individual output) during the final five rounds of the first part of the experiment was between $(i - 1)0$ and $i0$, and 0 otherwise, $i = 2, \dots, 7$. For example if $i = 5$,

then variable D_5 takes a value of 1 if the mean individual output of the group under investigation had a value between 40 $[(5 - 1)0]$ and 50 during the final five rounds of the first part of the experiment, and 0 otherwise.

D_{7+j} is a dummy variable taking a value of 1 if the plan used is plan j , and 0 otherwise, where $j = 1$ is the competitive team plan; $j = 2$ is the forcing contracts plan; $j = 3$ is the profit-sharing scheme; $j = 4$ is the gainsharing scheme, and

D_{11+k} is a dummy variable taking a value of 1 if the experience of the group (its mean individual output) during the first five rounds of the second part of the experiment was between $(i - 1)0$ and $i0$, and 0 otherwise, $i = 1, 2, \dots, 7, k = 1, 2, \dots, 7$.

The dummy variables and the regressions run are presented in Table 3, and the results of the regressions are presented in Tables 4–6.

Regression 1 looks at the set of non-revenue-sharing experiments which were run in Phase I where they were run after participation in a revenue-sharing experiment. The left-hand variable to be explained is the mean effort level of groups in the first five rounds of the non-revenue-sharing experiment, and two factors are used for this explanation: the mean effort levels experienced by these groups in the last five rounds of the revenue-sharing experiments and the current plan being used. In essence we are trying to separate the influence that experience has (as captured by mean effort levels in the last five rounds in a revenue-sharing experiment) from the influence of the incentives incorporated in the plan.

As we can see from Table 4, with the exception of the profit-sharing dummy (D_{10}), all plan and experience variables were significant variables in the regression. In addition, we can see from the negative coefficients in front of all experience variables that progressively worse experiences in the first revenue-sharing experiment lower mean output in the first five rounds of the subsequent non-revenue-sharing experiments. With respect to the plan, it appears that gainsharing and profit sharing are the worst plans; but performance increases with the introduction of either forcing contracts or competitive teams. (Note that the coefficient of the competitive team plan variable is significantly different from that of profit

TABLE 4—DUMMY VARIABLE REGRESSION: NON-REVENUE-SHARING EXPERIMENTS (PHASE-I EXPERIMENTS)

Variable	Coefficient	Standard error	<i>t</i>	<i>P</i> > <i>t</i>	95-percent confidence interval
Constant	39.422	4.488	8.738	0.000	30.300, 48.544
Experience					
<i>D</i> ₁	-19.002	6.502	-2.922	0.006	-32.216, -5.788
<i>D</i> ₂	-14.145	4.053	-3.490	0.001	-22.382, -5.908
<i>D</i> ₃	-12.142	4.413	-2.751	0.009	-21.111, -3.173
Plan					
<i>D</i> ₈	14.479	3.948	3.667	0.000	6.454, 22.503
<i>D</i> ₉	9.297	3.789	2.454	0.019	1.596, 16.998
<i>D</i> ₁₀	-4.407	4.595	-0.959	0.344	-13.7476, 4.931

Notes:

Number of observations = 41

Prob > F = 0.000

F(11,29) = 6.92

*R*² = 0.5498.Adjusted *R*² = 0.4703

sharing and implicitly different from gain-sharing as well.) Finally, it is interesting to see that with respect to output in the first five rounds of the non-revenue-sharing experiments, forcing contracts appears to be indistinguishable from competitive teams (at least as reflected in the confidence intervals about the estimated coefficients). This will not be the case at the end of the non-revenue-sharing experiments, as we will see below.

When we look at Phase II (where non-revenue-sharing experiments were run first followed by revenue sharing) and conduct Regression 2 (note, of course, that here plan refers to the plan used in the first, and not the second, part of the experiment where all groups engaged in revenue sharing), we find (see Table 5) that experience in the last five rounds of non-revenue-sharing experiments has no significant effect on how subjects start their revenue-sharing experiment, nor does the actual experiment they participated in during that first experiment. In short, these results suggest that performance of groups in a revenue-sharing experiment is independent of history—subjects tend to shirk no matter what mechanism they previously participated in and no matter what their experience with that plan was.

While these results above pertain to behavior during the first five rounds of the second experiment run, we also might be interested in how that same group ended up its experience with each other during the last five rounds of the second experiment. In Regression 3 the

mean group effort during the last five rounds of the non-revenue-sharing experiment was explained on the basis of three factors: the plan or incentive mechanism used (plan), the experience of the group (individual effort level) during the last five rounds of the first revenue-sharing experiment (Experience I), and the experience of the group (mean individual level) during the first five rounds of the second experiment (Experience II). The results of these regressions are presented in Table 6.

Note that when we have two experience variables, the variable Experience I becomes insignificant. In other words, behavior at the end of the second experiment appears only to be a function of how a group started that experiment and the actual plan under which it is functioning. However, such a conclusion is deceiving since we already know from Table 5 that the variables Experience I and Experience II are correlated. Hence, Experience I and Experience II are colinear variables with Experience I having at least an indirect effect on group performance at the end of the second experiment. Further, notice that it is particularly good experience in the first five rounds of the second experiment (*D*₁₀, the dropped dummy variable) that is a critical determinant of performance at the end of the second experiment. In fact, all other experiences significantly lower the mean effort level of groups during the last five rounds of the non-revenue-sharing experiment. Finally notice that the forcing contracts plan (*D*₉) is the only plan with a significant negative coefficient in ex-

TABLE 5—DUMMY VARIABLE REGRESSION: REVENUE-SHARING EXPERIMENTS (PHASE-II EXPERIMENTS)

Variable	Coefficient	Standard error	<i>t</i>	<i>P</i> > <i>t</i>	95-percent confidence interval
Constant	41.512	10.453	3.971	0.000	20.100, 62.932
Experience					
<i>D</i> ₁	-12.390	10.868	-1.140	0.264	-34.654, 9.8738
<i>D</i> ₂	-12.820	10.625	-1.207	0.238	-34.585, 9.9458
<i>D</i> ₃	-13.721	10.475	-1.310	0.201	-35.180, 7.7363
<i>D</i> ₄	-12.386	9.945	-1.245	0.223	-32.758, 7.9858
<i>D</i> ₅	-5.631	9.165	-0.614	0.544	-24.406, 13.144
<i>D</i> ₆	-8.120	12.08	-0.672	0.507	-32.875, 16.635
Plan*					
<i>D</i> ₈	-4.262	6.019	-0.708	0.485	-16.593, 8.068
<i>D</i> ₉	2.731	3.787	0.721	0.477	-5.026, 10.488

Notes:

Number of observations = 37 Prob > F = 0.857

Adjusted *R*² = -0.129 *R*² = 0.121.

F(8,28) = 0.480

*In this regression there is no gainsharing plan since that plan never preceded a revenue-sharing plan.

plaining mean output in the last five rounds, while competitive teams is the only plan with a positive coefficient (albeit insignificant). The fact that the forcing contract coefficient is negative indicates that the forcing contract plan is not stable in maintaining output levels throughout the history of the group interaction. In our experiments, the forcing contract regime consistently and substantially underperforms competitive teams in maintaining high output levels, even when the performance targets were significantly reduced and environmental uncertainty completely eliminated (making these plans less risky).⁵ Typically, after the group fails to reach the target sufficiently often, group output tumbles. These results cast doubt on the efficacy of exogenous targeting as a solution to the problem of group moral hazard.

Our results here are reminiscent of earlier results by Van Huyck et al. (1990, 1991). In those papers the authors examine whether subjects choose an equilibrium of a game on the basis of

⁵ In the set of experiments not reported here, we altered the parameters of the forcing contract regime, either reducing performance targets (to 240) for the entire group, or removing uncertainty from the group output function entirely. In these experiments, the comparative performance of competitive teams and forcing contracts remains qualitatively unchanged. A full discussion of these results can be found in Nalbantian and Schotter (1994).

the closest previous equilibrium chosen in games played immediately before. As in our work, they investigate whether a Pareto-dominant, yet risky or "vulnerable," equilibrium (see Nalbantian and Schotter [1994]; Schotter [1996]) will be chosen in a game if that group previously has had a successful experience playing a Pareto-dominant equilibrium of a less risky game. While their results are mixed, they find evidence that successful group-rational coordination in early games does not carry over to a continuation game, the Pareto-dominant equilibrium of which is risky. The logic behind our results is identical. On the other hand, when groups of laboratory workers have a common history of shirking, there is a lack of trust among them which causes them to avoid taking a chance on a risky, yet Pareto-dominant, equilibrium (this is what Figures 3–5 demonstrate). However, if they happen upon a very good common experience in their earlier interaction, then they build up a level of trust sufficient to allow them to take a chance in their subsequent interactions (at least during the first five periods of it). When the game they play has a dominant strategy equilibrium, like revenue sharing, then such a deductive characteristic overpowers any inductive selection principles inherent in the history of prior play of the group.⁶

⁶ For a closer look at the problem of trust and risk and their impact on worker productivity, see Schotter (1996).

TABLE 6—DUMMY VARIABLE REGRESSION: NON-REVENUE-SHARING EXPERIMENTS
LAST FIVE ROUNDS (PHASE-II EXPERIMENTS)

Variable	Coefficient	Standard error	<i>t</i>	<i>P</i> > <i>t</i>	95-percent confidence interval
Constant	73.171	13.338	5.486	0.000	45.890, 100.453
Experience I					
<i>D</i> ₁	-3.852	11.409	-0.338	0.738	-27.187, -19.482
<i>D</i> ₂	-4.930	6.837	-0.721	0.477	-18.913, 9.053
<i>D</i> ₃	-7.524	6.700	-1.123	0.271	-211.228, 6.180
Plan					
<i>D</i> ₈	7.798	7.400	1.054	0.301	-7.337, 22.934
<i>D</i> ₉	-16.740	5.428	-3.084	0.004	-27.843, -5.637
<i>D</i> ₁₀	-7.524	6.700	-1.123	0.271	-21.228, 6.180
Experience II					
<i>D</i> ₁₃	-52.150	14.531	-3.589	0.001	-81.870, -22.429
<i>D</i> ₁₄	-46.018	11.368	-4.048	0.000	-69.268, -22.767
<i>D</i> ₁₅	-38.053	11.754	-3.237	0.003	-62.092, -14.013
<i>D</i> ₁₆	-36.628	11.927	-3.071	0.005	-61.024, -12.233
<i>D</i> ₁₇	-52.038	14.461	-3.598	0.001	-81.616, -22.461

Notes:

Number of observations = 41 Prob > F = 0.001
 F(11,29) = 5.38 R² = 0.67154.
 Adjusted R² = 0.544

Finally, the impact of history on behavior is also pointed out by Isaac et al. (1991), where they survey a group of articles indicating that historical events may influence a person's perceptions of what is a fair outcome.

C. Observation 3: A Little Competition Goes a Long, Long Way

There really are two criteria one might want to use to evaluate any particular group incentive mechanism. First, one obviously would want to see what mean effort levels this mechanism defines. This is the aim of most principal-agent analyses. However, a corporate manager also might be interested in how reliable any given mechanism is in generating these high mean outputs. For example, say that a corporation has a number of identical plants situated across the country each producing identical products. At its disposal are two group incentive plans, Plan A and Plan B. Say that if Plan A were instituted in each of the corporation's plants, its mean aggregate output would be greater than Plan B's, but Plan B has less plant-to-plant variance attached to it. In other words, Plan B is a reliable producer of reasonably good outcomes, while Plan A has a higher mean but also a higher variance. Which plan one uses

obviously will depend on one's attitude toward risk. However, if one plan dominates the other in the sense of generating both higher mean outputs as well as smaller group-to-group variations, then clearly it should be chosen.

Observation 3 states that on these criteria the competitive teams mechanism practically dominates all other mechanisms tested, except for the fact that profit sharing and revenue sharing, while having significantly lower means, also have smaller variances around that mean than does the competitive team mechanism. When compared to forcing contracts or gainsharing, however, the competitive teams mechanism elicits both a higher mean effort level during the last five rounds of any experiment and a smaller across-group variance of effort around that mean.

Our support for Observation 3 is presented in Figures 9 and 10, which show the efficiency frontier for all of the mechanisms we have investigated in mean-variance space. In other words, we look at the mean group plan for each group during the last five rounds of each Phase-I and Phase-II experiment and the group-to-group variance across these groups in these same periods. Each point in the mean-variance space represents the mean-variance configuration for a specific plan. All vectors

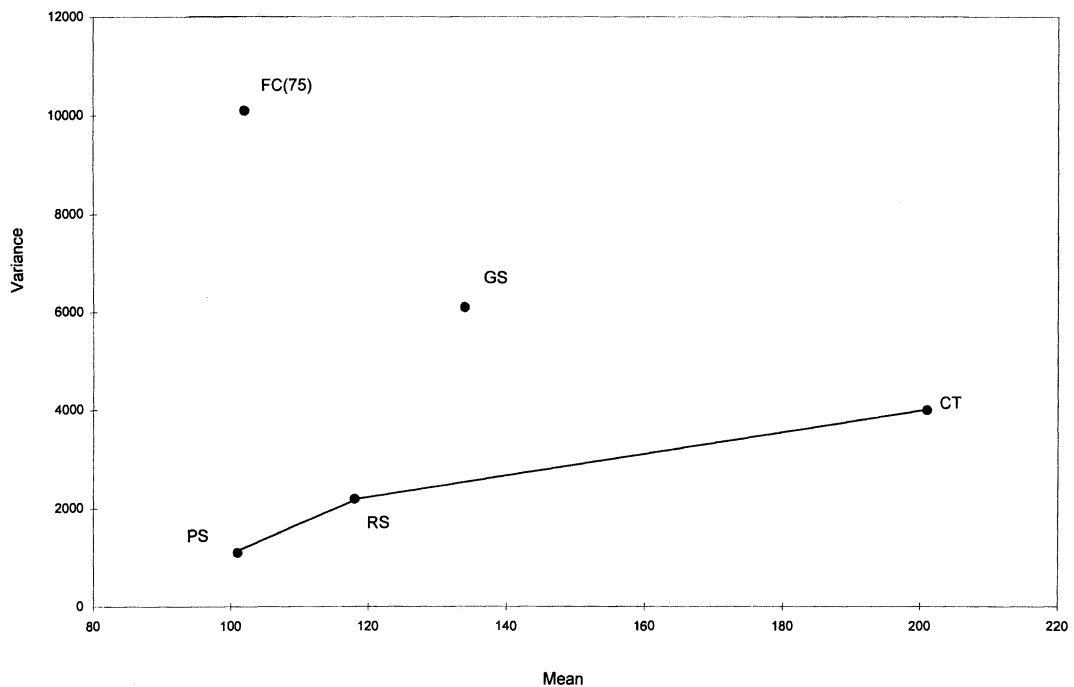


FIGURE 9. MEAN-VARIANCE PLOTS OF GROUP OUTPUT (PHASE I)

on the boundary of this set are connected. Note also that the competitive teams mechanism dominates all mechanisms except for profit sharing and revenue sharing. Further, since the mean of the competitive team experiment is so much higher than that of either profit sharing or revenue sharing, it is hard to conceive of a degree of risk aversion that would lead a corporate planner to actually prefer profit sharing over competitive teams. (Note, of course, that any risk-neutral corporate planner would prefer the competitive teams mechanism over all others.)⁷ It is in this sense that we claim that a little competition goes a long way.

⁷ Note that while the forcing contract (75) and competitive teams formulae have identical Pareto-optimal equilibria, profit sharing and revenue sharing do not. (Gainsharing has no predictable equilibrium outcome.) Hence, we should not expect profit sharing or revenue sharing to outperform these other mechanisms. However, since forcing contract (75) is, in essence, a profit-sharing mechanism with a Pareto-optimal target, comparing forcing contract (75) with competitive teams is equivalent to comparing a competitive teams mechanism to a profit-

D. *Observation 4: Monitoring Works But Is Costly*

As principal-agent theory tells us, if monitoring is possible it becomes quite easy to elicit optimal levels of effort from workers simply by monitoring them and firing them if they are caught shirking. When only imperfect monitoring is possible or monitoring is so expensive that workers can only be checked sporadically, the cost-effectiveness of such monitoring schemes is called into question. What principal-agent theory does not tell us is how sensitive workers will be to the detection probability of shirking. For instance, will even minor detection probabilities lead to high effort levels? Do workers misestimate the detection probability or suffer from some type of probability bias as is evidenced in other,

sharing mechanism. Also, revenue sharing is added to illustrate the mean-variance properties of a non-target-based scheme (remember, of course, that no revenue sharing, Pareto-optimal equilibrium exists).

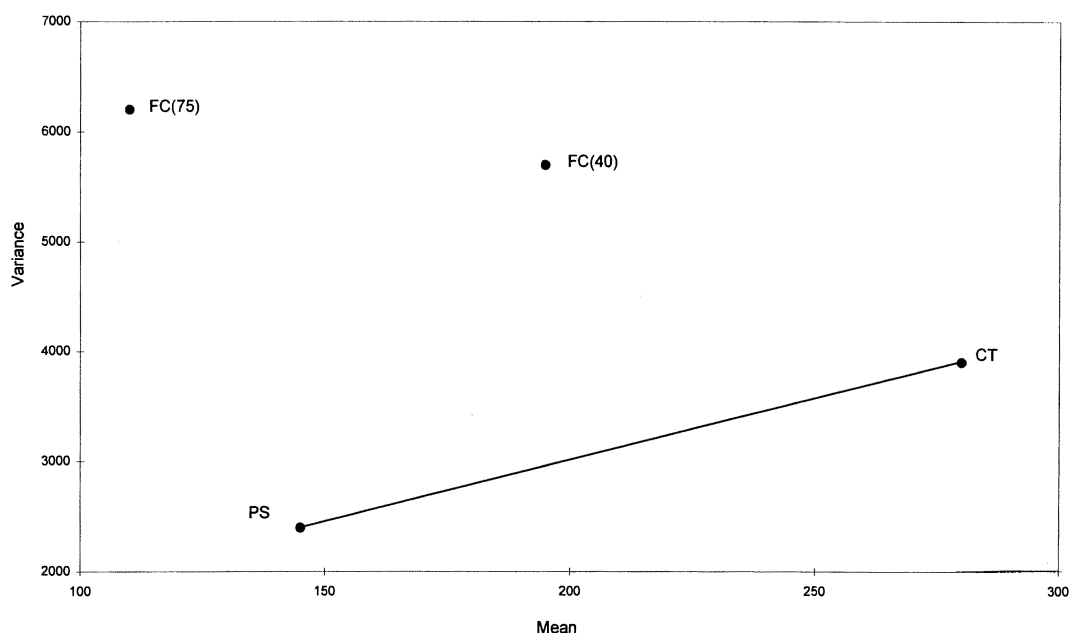


FIGURE 10. MEAN-VARIANCE PLOTS OF GROUP OUTPUT (PHASE II)

decision-under-uncertainty experiments, which leads them to consistently underestimate the probability of being caught? (See the survey by Colin Camerer [1995] for other instances of probability bias.)

Clearly these are questions that must be answered before we can suggest the relative superiority of monitoring schemes in corporations. Our experimental design furnishes data which give us some insight into this question. For instance, in Figures 11 and 12 we see the median effort level of subjects in our monitoring (0.70) and monitoring (0.30) experiments run in Phase I and II [i.e., in Phase I we ran our monitoring (0.70) experiment first and then our monitoring (0.30) experiment, while in Phase II the process was reversed]]. Clearly there is a dramatic difference between the median effort levels of subjects when being monitored with a 0.70 probability and a 0.30 probability. This is, of course, to be expected since the optimal response of subjects to a 0.70 monitoring probability is to choose a Pareto-optimal effort level, while the optimal response to a 0.30 monitoring probability is to shirk. This is seen in Figures 11 and 12, which show that groups functioning under a 0.70

monitoring probability choose Pareto-optimal effort levels as a median, while the 0.30 probability groups choose efforts the median of which involves almost complete shirking. Furthermore, while high detection levels (0.70) lead to consistently high effort levels whether the experiment was run before or after a low-detection experiment, low detection levels (0.30) lead to quite different types of behavior in experiments run before and after high-detection experiments. This is clearly seen in Figure 12, where the median effort level for the low-detection group is practically zero in all periods when the experiment is performed in Phase I (after the 0.70 detection experiment). When it is performed in Phase II (before the 0.70 experiment), the results are quite different. What is striking is that the predictions of the theory seem to fail when subjects are not experienced and when they are subject to the low 0.30 monitoring probability (median effort levels are periodically above the zero effort levels). Obviously, when groups are used to high-detection probabilities, a drop to low probabilities seems to lead them to reevaluate their effort choices and lower them. Again, history matters.

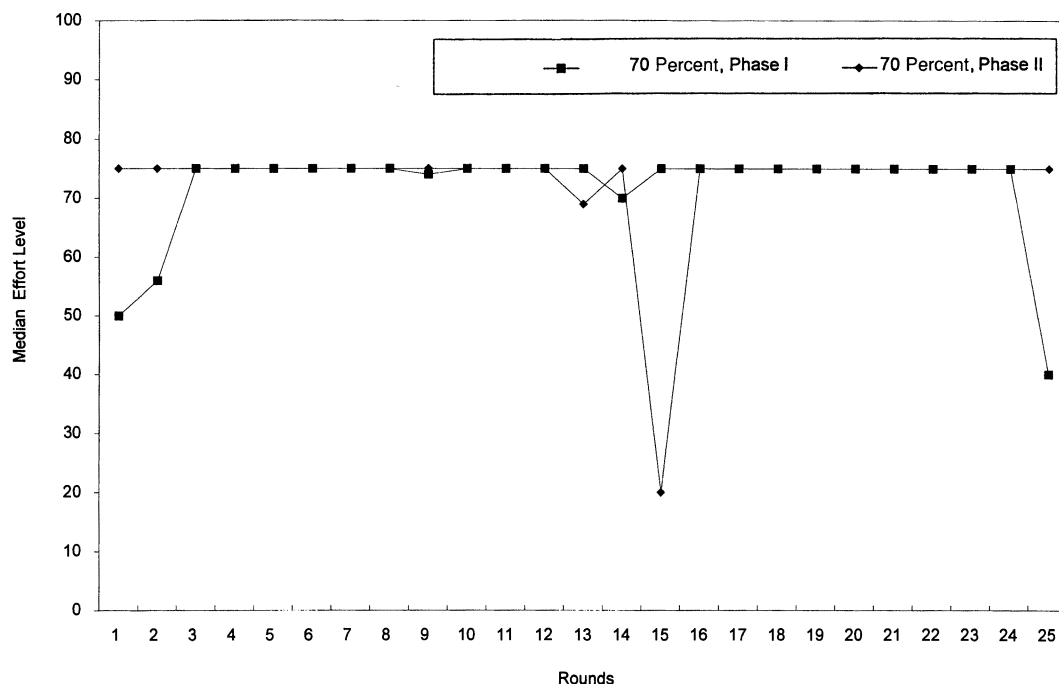


FIGURE 11. MONITORING (70-PERCENT PROBABILITY)—MEDIAN EFFORT LEVELS

V. Conclusions

This paper has attempted to take a first step on the road to adding some empirical meat to the skeleton created by theorists working on the problem of group incentives and productivity. In our experiments we have uncovered a number of factors which we think are probably important for the proper design of group incentive mechanisms. Most important among these findings are the observations which follow. First, we have found that the history of a group and its performance in the past is an important predictor of how that group will perform when a new incentive program is introduced. In addition, we have discovered evidence that one effective way to increase group effort is to introduce some within-firm competition between work units performing the same task—setting up an intrafirm team tournament. Targets established endogenously on the basis of relative performance do better at stimulating group output than those that are externally stipulated.⁸

⁸ Of course, we recognize that where there is substantial interdependence across work groups, the use of relative

Finally, although we do not report these results here (see Nalbantian and Schotter [1994]; Schotter [1996]), our findings indicate that it is not sufficient to expect a mechanism to implement Pareto-optimal outcomes as Nash equilibria without taking into account the out-of-equilibrium properties of the mechanism. If those optimal outcomes actually are to be achieved, it is necessary that those equilibria be relatively riskless or not “vulnerable” to slight mistakes by one’s colleagues. Mechanisms which attach considerable risk to selecting the efficient equilibrium ultimately may lead economic agents to opt out of the mechanism and play it safe by shirking as was the case in our forcing contracts plan.

While these findings strike us as interesting, we are well aware of their limitations. To begin with, our experiments provide at best a bare-bones economist’s view of the incentive problem. They characterize productive

performance evaluation can undermine incentives for cooperation and reduce organizational productivity.

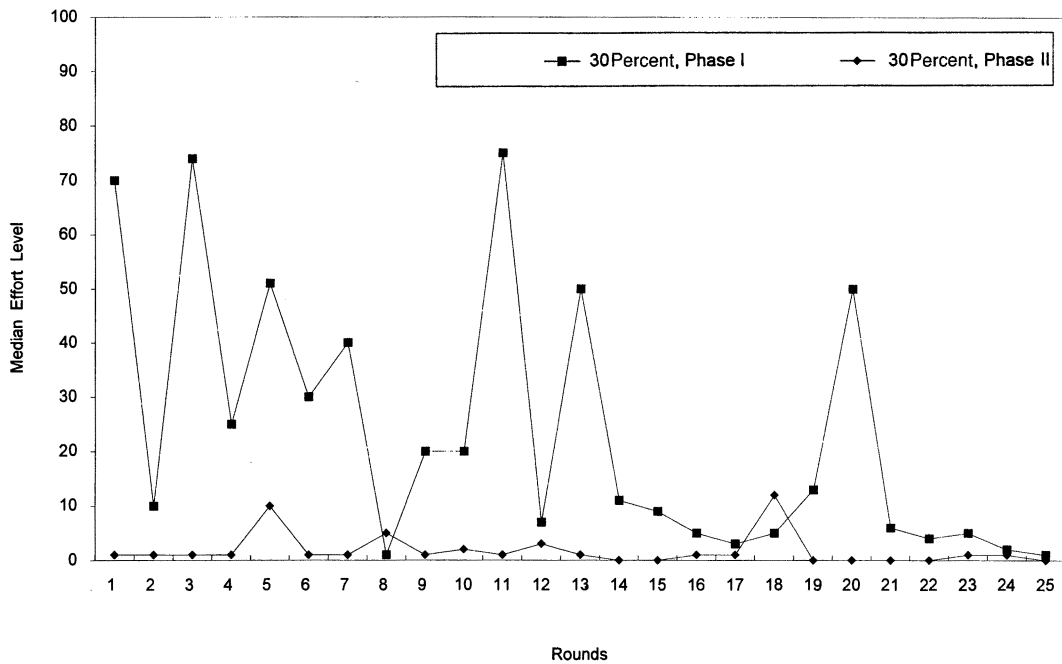


FIGURE 12. MONITORING (30-PERCENT PROBABILITY)—MEDIAN EFFORT LEVELS

performance as the outcome of a noncooperative game and are concerned solely with the incentive properties of the various reward formulae as the explanation of behavior. Psychologists, compensation practitioners, and others certainly would protest that life is far different in the workplace than in our experiments. The character of interactions among workers is considerably more complex than that presumed here. For one thing, work could just as well be modeled as a cooperative game in which workers communicate with each other and implicitly establish work norms which they then enforce upon each other. Results from Dawes et al. (1977) indicate that communication between subjects dramatically affects their performance in public goods experiments. Similarly, while managers may not be able to monitor workers effectively, workers themselves may be in a better position to do so. And they would have more incentive to perform that function in a system where their rewards depend on co-worker performance than in a system where they depend on one's own performance alone (Nalbantian, 1987; David I. Levine and Laura D'Andrea

Tyson, 1990; Eugene Kandel and Edward P. Lazear, 1992). A full and proper experimental design should allow this factor to come into play.

These are all valid points, but we do not feel that they reduce the significance of what we have uncovered here. Our intent in these experiments was to see how far the orthodox economic model of group incentives and the "corrective" reward formulae deduced therefrom can take us in explaining behavior. As previously noted, substantial evidence from the field does indeed suggest that observed behavior under group incentives is often at variance from that predicted by standard theory. Certain other factors clearly are central to the operation of group incentives and, therefore, should be incorporated in experimental treatments of the problem. We intend to do so in our future research. Still, in conducting these experiments we have learned important lessons about the properties of prototypical group incentive formulae—findings that transcend the particular context in which they were revealed. We have established an economic baseline which helps determine how much

column 2), the computer will add them up. We will call the resulting number the group total. The computer will then randomly choose a number between -40 and $+40$ and add it to the group total. When we say "randomly," we mean that each number in the interval -40 to $+40$ has an equal chance of being chosen. Hence, the chance of -30 being chosen is equal to the chance of $+15$ being chosen, which in turn is equal to the chance of $+5$ being chosen, and so on. Finally, the sum of this random number and the group total will be multiplied by the number 1.5 (francs) to get what we call group revenue, which will appear in column 3 on your screen as "Group Rev." For example, say that the decision numbers of the six members of your group are z_1 , z_2 , z_3 , z_4 , z_5 , and z_6 , where z_1 is the decision number of subject #1, z_2 is the decision number of subject #2, and so on. Further, suppose that the random number generated by the computer is $+5$. Then the group total would be: $(z_1 + z_2 + z_3 + z_4 + z_5 + z_6 + 5)$ and the group revenue would be $1.5(z_1 + z_2 + z_3 + z_4 + z_5 + z_6 + 5)$. As you can see, group revenue will thus reflect both the choices of each group member regarding his/her decision number **and** the realization of the random number, namely, pure chance.

Group revenue (group total) is the basis of your individual payment. Specifically, in each round of the experiment your group will be given a target group revenue of 112.5 francs. (Note: this corresponds to a group total of 75, i.e., $75 \times 1.5 = 112.5$.) If your group revenue turns out to be less than 112.5, your payment for the round will be zero. If your group revenue precisely equals 112.5, your payment for the round will be the fixed amount of 18.75 francs. On the other hand, if your group revenue exceeds 112.5 francs, your personal payment will be the sum of 18.75 francs and one-sixth of the difference between your group revenue and 112.5 francs. In other words, in addition to the fixed amount, 18.75 francs, you personally will be paid one-sixth of the excess of actual group revenue over the target group revenue. For example, say that your group revenue is 172.5 francs, which exceeds the target of 112.5 by 60 francs. Your payment for this round would then be $18.75 + 1/6(172.5 - 112.5) = 28.75$ francs.

Clearly, above the 112.5-franc threshold or "group target," the larger is group revenue, the greater your payment will be—though as you will see, you will have to deduct from your payment the cost associated with your decision number. Below the group target, your payment is a fixed amount (0 francs) independent of group revenue.

The group target of 112.5 francs is indicated in column 4 on your screen. Your payment for each round of the experiment will be calculated by the computer and appear in column 5 on your screen.

How Your Earnings Are Determined

Your payoff or earnings in any round will equal the payment you receive, as described above, **minus** the cost of your decision number. Decision costs are presented in the cost schedule table. You will note that for each decision number you might choose over the range 0 to 100, there is an associated cost to be incurred. You can read your cost table by looking down the first column and finding the decision number you are contemplating. The second column will then inform you what it will cost you to choose that decision number. For example, a decision number of 25 has an associated cost of 6.75 francs, while the decision number 50 has a cost of 25 francs. Several important features of this cost schedule are evident in this example and are especially noteworthy. First, the larger the decision number, the higher the cost you must incur. Second, the cost of decision numbers increases *at an increasing rate*. Hence, the cost of choosing decision number 50 is more than twice the cost of choosing 25; The cost of choosing 100 is more than twice the cost of choosing 50. You can verify this characteristic of costs of decision numbers by considering other examples from the cost schedule.

The cost of the decision number you choose will be deducted from the payment you are due in that round to determine your actual earnings for the round. Again, earnings = payment – decision cost. The cost of your decision number for each round will appear in column 6 on your screen.

To illustrate how your earnings will be determined, suppose that group revenue in round 1 of the experiment is calculated at 200 francs

and that the decision number you selected in that round was 40. Since 200 is greater than 112.5 (the group target), your payment then would be calculated as: $18.75 + 1/6[200 - 112.5] = 33.33$ francs. From the cost schedule table we find that the cost of decision number 40 which you chose is 16 francs. Therefore, your earnings for round 1 would be: $33.33 - 16 = 17.33$ francs. Suppose, on the other hand, that group revenue in a given round is 112.5 francs, and that your decision number is again 40. Since the group target is precisely attained, your payment will be 18.75 from which your decision cost must be deducted; your earnings are then calculated as $18.75 - 16 = 2.75$ francs. Finally, suppose your group revenue is 110, while your decision number remains 40. Since the group target has not been achieved, your payment is 0 francs. Thus your earnings for this round would be $0 - 16$ francs = -16 francs. (Negative earnings will be deducted from your accumulated earnings at the end of the experiment.)

Your earnings, or payoff, for the round are calculated by the computer and appear in column 7 on your screen.

Final Payoffs

Your final payoff in the experiment will be equal to the sum of the francs received over the 25 rounds of the experiment. Each franc will be converted at the rate of 1 franc = .71 cents. In addition to this payoff, you will receive a fixed payoff of \$3.00 just for showing up at the experiment.

REFERENCES

- Banks, Jeffrey S.; Plott, Charles R. and Porter, David P.** "An Experimental Analysis of Unanimity in Public Goods Provision Mechanisms." *Review of Economic Studies*, April 1988, 55(2), pp. 301-22.
- Blasi, Joseph R.** *Employee ownership: Revolution or ripoff?* Cambridge, MA: Ballinger, 1988.
- Blasi, Joseph R. and Kruse, Douglas L.** *The new owners*. New York: Harper Business, 1991.
- Blinder, Alan S.** *Paying for productivity: A look at the evidence*. Washington, DC: Brookings Institution, 1990.
- Bull, Clive; Schotter, Andrew and Weigelt, Keith.** "Tournaments and Piece Rate: An Experimental Study." *Journal of Political Economy*, February 1987, 95(1), pp. 1-33.
- Cable, John and Wilson, Nicholas.** "Profit-Sharing and Productivity: An Analysis of the UK Engineering Firms." *Economic Journal*, June 1989, 99(396), pp. 366-75.
- Calvo, Guillermo A.** "The Economics of Supervision," in Haig R. Nalbantian, ed., *Incentives, cooperation, and risk sharing: Economic and psychological perspectives on employment contracts*. Totawa, NJ: Rowman & Littlefield, 1987, pp. 87-103.
- Calvo, Guillermo A. and Wellisz, Stanislaw H.** "Supervision, Loss of Control, and the Optimum Size of the Firm." *Journal of Political Economy*, October 1978, 86(5), pp. 943-52.
- Camerer, Colin.** "Individual Decision-Making," in John H. Kagel and Alvin E. Roth, eds., *Handbook of experimental economics*. Princeton, NJ: Princeton University Press, 1995.
- Conte, Michael and Svejnar, Jan.** "Employee Share Ownership, Profit Sharing and Enterprise Performance." Mimeo, ILR Press-Cornell University, 1989.
- Dawes, Robyn M.; McTavish, Jeanne and Shaklee, Harriet.** "Behavior, Communication, and Assumptions about other People's Behavior in a Commons Dilemma Situation." *Journal of Personality and Social Psychology*, January 1977, 35(1), pp. 1-11.
- Dawes, Robyn M.; Orbell, John M.; Simmons, Randy T. and van de Kragt, Alphons J. C.** "Organizing Groups for Collective Action." *American Political Science Review*, December 1986, 80(4), pp. 1171-85.
- Fein, Mitchell.** "Improved Productivity Through Worker Involvement." Mimeo, Mitchell Fein, Inc., 1982.
- FitzRoy, Felix and Kraft, Kornelius.** "Profitability and Profit-Sharing." *Journal of Industrial Economics*, December 1986, 35(2), pp. 113-30.
- _____. "Cooperation, Productivity and Profit Sharing." *Quarterly Journal of Economics*, February 1987, 102(1), pp. 23-35.
- Graham-Moore, Brian E. and Ross, Timothy L.** *Productivity gainsharing: How employee incentive programs can improve company performance*. Englewood Cliffs, NJ: Prentice-Hall, 1983.

- Holmstrom, Bengt R.** "Moral Hazard in Teams." *Bell Journal of Economics*, Autumn 1982, 13(2), pp. 324-40.
- Howard, Bion B. and Dietz, Peter O.** *A study of the financial significance of profit sharing*. Chicago: Council of Profit Sharing Industries, 1969.
- Huselid, Mark A.** "The Impact of Human Resource Management Practices on Turnover, Productivity and Corporate Financial Performance." *Academy of Management Journal*, June 1995, 38(3), pp. 635-72.
- Ichniowski, Casey and Shaw, Kathryn.** "Old Dogs and New Tricks: Determinants of the Adoption of Productivity-Enhancing Work Practices." *Brookings Papers on Economic Activity: Microeconomics*, 1995, pp. 1-55.
- Ichniowski, Casey; Shaw, Kathryn and Prenzushi, Giovanna.** "The Effects of Human Resource Practices on Productivity." National Bureau of Economic Research (Cambridge, MA) Working Paper No. 5333, 1996.
- Isaac, R. Mark; Mathieu, Deborah and Zajac, Edward E.** "Institutional Framing and Perceptions of Fairness." *Constitutional Political Economy*, Fall 1991, 2(3), pp. 329-70.
- Isaac, R. Mark and Walker, James M.** "Group Size Effects in Public Goods Provision: The Voluntary Contribution Mechanism." *Quarterly Journal of Economics*, February 1988, 103(1), pp. 179-99.
- Kandel, Eugene and Lazear, Edward P.** "Peer Pressure and Partnership." *Journal of Political Economy*, August 1992, 100(4), pp. 801-18.
- Kruse, Douglas L.** "Profit Sharing and Productivity: Microeconomic Evidence from the United States." *Economic Journal*, January 1992, 102(410), pp. 24-36.
- Ledyard, John O.** "Public Goods: A Survey of Experimental Research," in John H. Kagel and Alvin E. Roth, eds., *Handbook of experimental economics*. Princeton, NJ: Princeton University Press, 1995.
- Levine, David I. and Tyson, Laura D'Andrea.** "Participation, Productivity and the Firm's Environment," in Alan S. Blinder, ed., *Paying for productivity: A look at the evidence*. Washington, DC: Brookings Institution, 1990, pp. 183-237.
- Marwell, Gerald and Ames, Ruth E.** "Experiments on the Provision of Public Goods II: Provision Points, Stakes, Experience, and the Free-Rider Problem." *American Journal of Sociology*, January 1980, 85(4), pp. 926-37.
- McAdams, Jerry L. and Hawk, Elizabeth J.** *Organizational performance and rewards: 663 experiences in making the link*. Scottsdale, AZ: Consortium for Alternate Reward Strategies Research, American Compensation Association, 1994.
- Metzger, Bertram L.** *Profit sharing in 38 large companies, I & II*. Evanston, IL: Profit Sharing Research Foundation, 1975.
- Metzger, Bertram L. and Colletti, Jerome A.** *Does profit sharing pay?* Evanston, IL: Profit Sharing Research Foundation, 1971.
- Nalbantian, Haig R.** "Incentive Compensation in Perspective," in Haig R. Nalbantian, ed., *Incentives, cooperation, and risk sharing*. Totowa, NJ: Rowman & Littlefield, 1987, pp. 3-43.
- Nalbantian, Haig R. and Schotter, Andrew.** "Productivity Under Group Incentives: An Experimental Study." Economic Research Report 94-04, C.V. Starr Center for Applied Economics, New York University, January 1994.
- National Commission on Productivity and Work Quality.** *A national policy for productivity improvement*. Washington, DC: U.S. Government Printing Office, 1975.
- New York Stock Exchange Office of Economic Research.** *People and productivity*. New York: New York Stock Exchange, 1982.
- O'Dell, Carla and McAdams, Jerry L.** *People, performance and pay*. Houston, TX: American Productivity Center, 1987.
- Palfrey, Thomas and Rosenthal, Howard.** "Testing Game Theoretic Models of Free-Riding: New Evidence of Probability Bias and Learning," in T. Palfrey, ed., *Laboratory research in political economy*. Ann Arbor, MI: University of Michigan Press, 1991, pp. 239-68.
- Pearce, Jone L.; Stevenson, William B. and Perry, James L.** "Managerial Compensation Based on Organizational Performance." *Academy of Management Journal*, June 1985, 28(2), pp. 261-78.
- Schotter, Andrew.** "Worker Trust and System Vulnerability and the Performance of Work Groups." Economic Research Report 96-32, C. V. Starr Center for Applied Economics, New York University, September 1996.

- Schotter, Andrew and Weigelt, Keith.** "Behavioral Consequences of Corporate Incentives and Long-Term Bonuses: An Experimental Study." *Management Science*, September 1992a, 38(9), pp. 1280-98.
- _____. "Asymmetric Tournaments, Equal Opportunity Laws, and Affirmative Action: Some Experimental Results." *Quarterly Journal of Economics*, May 1992b, 107(2), pp. 511-39.
- Smith, Vernon L.** "The Principle of Unanimity and Voluntary Consent in Social Choice." *Journal of Political Economy*, December 1977, 85(6), pp. 1125-39.
- U.S. Department of Labor.** *High performance work practices and firm performance*. Washington, DC: U.S. Government Printing Office, 1993.
- U.S. General Accounting Office.** *Productivity sharing programs: Can they contribute to improvement?* Washington, DC: U.S. Government Printing Office, 1981.
- Van Huyck, John B.; Battalio, Raymond C. and Beil, Richard O.** "Tacit Coordination Games, Strategic Uncertainty, and Coordination Failure." *American Economic Review*, March 1990, 80(1), pp. 234-48.
- _____. "Strategic Uncertainty, Equilibrium Selection, and Coordination Failure in Average Opinion Games." *Quarterly Journal of Economics*, August 1991, 106(3), pp. 885-910.
- Wagner III, John A.; Rubin, Paul A. and Callahan, Thomas J.** "Incentive Payment and Non-Managerial Productivity: An Interrupted Time Series Analysis of Magnitude and Trend." *Organizational Behavior and Human Decision Processes*, August 1988, 42(1), pp. 47-74.
- Weiss, Andrew.** "Incentives and Worker Behavior: Some Evidence," in Haig R. Nalbantian, ed., *Incentives, cooperation, and risk sharing*. Totowa, NJ: Rowman & Littlefield, 1987, pp. 137-50.
- Weitzman, Martin L. and Kruse, Douglas L.** "Profit Sharing and Productivity," in Alan S. Blinder, ed., *Paying for productivity: A look at the evidence*. Washington, DC: Brookings Institution, 1990, pp. 95-140.