

DISCUSSION PAPER SERIES

No. 2720

PERFORMANCE INCENTIVES WITH AWARD CONSTRAINTS

Pascal Courty and Gerald Marschke

INDUSTRIAL ORGANIZATION



Centre for **E**conomic **P**olicy **R**esearch

www.cepr.org

Available online at:

www.cepr.org/pubs/dps/DP2720.asp

PERFORMANCE INCENTIVES WITH AWARD CONSTRAINTS

Pascal Courty, London Business School and CEPR
Gerald Marschke, State University of New York

Discussion Paper No. 2720
March 2001

Centre for Economic Policy Research
90–98 Goswell Rd, London EC1V 7RR, UK
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999
Email: cepr@cepr.org, Website: www.cepr.org

This Discussion Paper is issued under the auspices of the Centre's research programme in **Industrial Organization**. Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as a private educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions. Institutional (core) finance for the Centre has been provided through major grants from the Economic and Social Research Council, under which an ESRC Resource Centre operates within CEPR; the Esmée Fairbairn Charitable Trust; and the Bank of England. These organizations do not give prior review to the Centre's publications, nor do they necessarily endorse the views expressed therein.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

Copyright: Pascal Courty and Gerald Marschke

CEPR Discussion Paper No. 2720

March 2001

ABSTRACT

Performance Incentives with Award Constraints*

This Paper studies the provision of incentives in a large US training organization, which is divided into about 50 independent pools of training agencies. The number and the size of the agencies within each pool vary greatly. Each pool distributes performance incentive awards to the training agencies it supervises, subject to two constraints: the awards cannot be negative and the sum of the awards cannot exceed an award budget. We characterize the optimal award function and derive simple predictions about how award prizes should depend on the number of agencies, on their sizes, and on their performances. Our results indicate that the constraints on the award distribution bind and reduce the overall efficiency of the incentive system.

JEL Classification: H72, J33, L14

Keywords: fixed award budget, limited liability, performance incentive

Pascal Courty
London Business School
Sussex Place
Regent's Park
London NW1 4SA
UK
Tel: (44 20) 7262 5050 x3317
Fax: (44 20) 7402 0718
Email: pcourty@lbs.ac.uk

Gerald Marschke
Department of Economics and
Department of Public Administration
and Policy
State University of New York, Albany
BA-110
Albany, NY 12222
USA
Tel: (1 518) 442 5274
Fax: (1 518) 442 5298
Email: marschke@csc.albany.edu

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=134679

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=

*This Paper is a revision of a paper entitled 'Designing Optimal Award Functions: Theory and Application to a Government Organization' that was prepared for the NAS conference on 'Devising Incentives to Promote Human Capital'. We would like to thank James Heckman for stimulating our interest in the study of incentives in government organizations. We thank Margaret Meyer as well as seminar participants at UCL and participants at conferences organized by CEPR, CMPO, ESSET and NAS. All errors are exclusively the responsibility of the authors.

Submitted 22 November 2000

NON-TECHNICAL SUMMARY

An important contribution of economic theory is the paradigm of incentive provision under moral hazard. The stereotype model of moral hazard has a principal who contracts with an agent to supply effort. Principal–agent relationships include the owners of a firm and the firm’s manager, a landlord and a farmer, a homeowner and a builder, to name just a few examples. The principal does not observe the agent’s effort so it is impossible to reward the agent for their contribution to what the principal truly values. The principal only observes a measure of performance that imperfectly represents effort. The performance measure reflects, in addition to effort, external factors that are outside of the agent’s control. An important feature of these models is that to induce effort, the principal will have to reward high performances and punish low ones even though – given the incentive system – the difference in performance is due to pure chance. Any model of incentive provision (e.g. piece rate, tournament, and relative performance evaluation) is a variation along that theme.

A key implication common to these models is that the agent sometimes receives more than their contribution and at other times less. Although this assumption that punishments and rewards can vary more than contribution plays a central role in the theory of incentives, it is often violated in practice. For example, punishments are often bounded from below: the principal cannot punish the agent beyond some given point. Incentive theorists have devoted some attention to this case, known as limited liability.

Another constraint on incentive provision, which has received less attention in the literature, is that awards often cannot be greater than a fixed award pot. The incentive literature has focused almost exclusively on situations where there is a surplus to be shared (e.g. executives and stockholders share stock market value creation, peasants and landlords share crops, and firms and sales people share sales margins). In many contractual relationships, however, the agent’s contribution does not directly materialize as a surplus that can be objectively valued and shared. In these situations, organizations typically prefer to set aside an award pot rather than, for example, taking the risk of committing to a subjective award formula that may bankrupt the incentive system. When the incentive scheme has to be fully funded, the agents’ awards are bounded from above because the agent cannot receive more than the award pot. Tournaments are examples of fully funded award schemes.

The main purpose of this Paper is to investigate whether the limited liability and fully funded constraints lower the efficiency of incentives. From a theoretical point of view, there are good reasons to believe that they should. The simplest way to see that is to revisit the principal–agent model of performance incentives after introducing these constraints. As before, the

principal will try to stimulate effort by creating a reward gap between high and low levels of agent performance. Limited liability constraints, however, may restrict the ability to give awards that are low relative to the agent's contribution. Similarly, fully funded constraints may limit the principal's ability to give rewards that are large relative to the agent's contribution. This upper bound on the rewards together with the lower bound on punishment may reduce the maximum award gap and the possibility to efficiently provide performance incentives.

This Paper investigates the hypothesis that constraints on incentive awards lower their effectiveness, using evidence from a large US government organization that provides job training to the economically disadvantaged. This job training organization divides the US into approximately 640 non-overlapping, collectively exhaustive jurisdictions, each run by a single training agency. Because a training agency's budget is determined primarily by the density of the population of disadvantaged that live in its jurisdiction, budgets vary in size across training agencies. Each state supervises the set of training agencies, or local decision-makers, that are located within its boundaries. Training agencies make resource allocation decisions freely, but face a financially backed incentive system that is based on a set of well defined performance measures. Each state distributes budgetary awards from a fixed award pot (fully funded constraint) to provide incentives subject to the constraint that awards cannot be negative (limited liability constraint). In fact, training agencies are guaranteed a fixed budget and receive the awards on top of that budget.

In investigating whether these constraints matter, this Paper proceeds in two steps. First, we characterize the distortions that emerge under the fully funded and the limited liability constraints. We proceed by constructing a two-agent model that captures the feature of our empirical application that the agents manage budgets of different sizes. The feature that budget size differs gives us an additional degree of freedom in identifying the implications of our model in the data. The model predicts that the limited liability and the fully funded constraints should bind, thereby reducing the effectiveness of incentives. The model also predicts that these effects will be more pronounced in states that manage a more diverse set of budgets. We show that when the relative difference in agents' budgets reaches a given threshold, the large agent exerts inefficiently low levels of effort.

Second, we test if the predictions suggested by the optimal contract hold in the federal job training organization that is our case study. Our empirical strategy is to compare performance awards and performance outcomes across states that manage different pools of agents. The empirical analysis uncovers three findings. First, those agents that are small relative to their state average receive disproportionately larger awards. Second, we find mixed evidence that these smaller agents perform better. Third, we find some

evidence that performance outcomes are lower in states that are more heterogeneous.

The evidence is broadly consistent with the predictions of the model and suggests that it is more difficult to provide performance incentives in states that are more heterogeneous because the fully funded and limited liability constraints are more likely to bind in those states. Our analysis shows that the effectiveness of performance incentive depends on the financial constraints organizations face. The analysis also suggests that the sorting of agents into pools is an important step in the design of incentive systems.

1 Introduction

This paper studies the design of incentives in a large federal (U.S.) organization that provides job training to the economically disadvantaged. State boundaries segment the organization. Each state supervises the training agencies, or local decision makers, that are located within its boundaries. Training agencies are heterogeneous in the sense that they manage budgets of different sizes. Training agencies' budgets are determined primarily by the density of the population of disadvantaged that live in their jurisdictions.

Each state distributes an award pot to provide incentives to the pool of training agencies it oversees subject to two constraints. First, the awards cannot be negative. Training agencies are guaranteed a fixed budget and receive the awards on top of that budget. This constraint is similar to the limited liability constraint found in the incentive literature (Sappington, 1983). The second constraint is that the award function has to be fully-funded. By this, we mean that the sum of the rewards cannot be greater than a fixed award pot. Tournaments are examples of fully-funded awards.

The main purpose of this paper is to investigate whether the limited liability and fully-funded constraints matter. From a theoretical point of view, there are good reasons to believe that they should. To see that, recall that the driving force behind performance incentives is that the way the principal stimulates effort is by creating a reward gap between high and low levels of agent performance. Under moral hazard, this implies that the agent will sometimes receive less and other times more than its contribution. Limited liability constraints, however, restrict the ability to give less to the agent than its contribution. Similarly, fully-funded constraints limit the principal's ability to give rewards that are greater than the agent's contribution. This upper-bound on the rewards together with the lower-bound on punishment due to limited liability may reduce the maximum award gap and the possibility to efficiently provide performance incentives.

In investigating whether these constraints matter, this paper proceeds in two steps. The first step is to model the contractual features described above. The incentive literature has overlooked situations where fully funded and limited liability constraints interplay with the feature that agents are heterogeneous. The model asks three sets of questions.

The first set explores the relation between the agents' performances and their awards. What does the optimal incentive scheme look like? Should the awards be independent across agents as in a piece rate system or should the amount agents receive depend on the performance of other agents, as in a tournament incentive scheme? The second set of questions is specific to the feature that agents are heterogeneous. How does the optimal award function depend on the number of agents and on their relative sizes? Should awards be proportional to budget sizes? Or should smaller agents receive a disproportional fraction of the award? Third and most importantly, does the optimal contract achieve the efficient level of effort? Do the limited liability and fully funded constraints bind?

The model predicts that the limited liability and the fully-funded constraints should bind and reduce the effectiveness of incentives. This should be even more pronounced in states where the agents' sizes are more heterogeneous. We show that when agents are very heterogeneous, the smaller agents will typically exert inefficiently high level of efforts. We also derive the optimal incentive contract and characterize its properties. Some of these properties suggest simple predictions on how budget sizes, award amounts and performance outcomes should vary within and across states. We also find that the optimal award is characterized by group incentives. An agent's payoff is dependent on the performance of her peers even though their performances may not be statistically related. The reason for the optimality of group incentives here comes from the need to cross-subsidize awards in order to increase the award gap between high and low levels of performance.

The second step is to test if the predictions suggested by the optimal contract hold in the federal job training organization that is our case study. Our empirical strategy is to compare performance awards and performance outcomes across states that manage different pools of agents. The empirical analysis uncovers three findings. First, those agents that are small relative to their state average receive disproportionately larger awards. We also find some mixed evidence that they perform better. Second, performance awards depend on absolute performance outcomes but also on performance outcomes relative to other agents in the state. Third, we find some evidence that performance outcomes are lower in states that are more heterogeneous. The evidence is broadly consistent with

the predictions of the model. It suggests that it is more difficult to provide performance incentives in states that are more heterogeneous because the fully-funded and limited liability constraints are more costly in those states.

The theoretical part of this paper contributes to the contract literature. Following the early work of Lazear and Rosen (1981) on tournaments as a means to provide incentives, some authors have recently studied the specific problem of allocating fixed award pots among contestants (e.g. Krishna and Morgan (1998) and Moldovanu and Sela (1999)) but these work do not assume limited liability on the part of the agent. As mentioned above, Sappington as well as Demski et al. (1988) study the restriction imposed by limited liability constraints but in a framework where the agent receives some private information after contracting. More recently, Innes (1990), and Kim (1997) considered the contractual restrictions imposed under limited liability but in a single agent framework and without the fully funded constraint.

On the empirical side, this work belongs to the empirical literature on the provision of incentive in organizations. See Prendergast (1999) for a recent survey of that literature. Another way to interpret our results is as a test of whether government bureaucrats write contracts that are consistent with the optimal incentive contracts predicted by incentive theory. There is some evidence that firms design optimal incentive contracts (Prendergast reviews studies of bonus, relative performance, and tournament) but to our knowledge, no one has yet asked whether government organizations also do so.

The paper is organized as follows. The next Section summarizes the key characteristics of the incentive system we study in the empirical application. This will be the starting point to motivate the model which is presented in Section 3. Section 4 derives some implications but the proofs are deferred to the Appendix. Section 5 tests some of the model's implication in a large training organization and Section 6 concludes.

2 The JTPA Incentive System

The Job Training Partnership Act (JTPA) of 1982 created what was until the late 90's the largest federal employment and training program serving the disadvantaged.¹ The core of our empirical work focuses on fiscal years 1985 and 1986. In these years, the JTPA annual budget was approximately \$4 billion and it was serving nearly one million people. JTPA is highly decentralized: job training is carried out by more than 600 semi-autonomous sub-state training agencies. The JTPA bureaucracy is unusual for many reasons but one will be of special interest for this study: Instead of a rigid, comprehensive set of rules that regulate bureaucratic *conduct*, the JTPA organization is driven by a set of incentive systems that influence *outcomes*.²

JTPA gave the responsibility to individual states to design and administer the local incentive systems. There are 51 incentive systems in our data set corresponding to 50 individual states and the District of Columbia. Each incentive system rewards a pool of training agencies. In fiscal years 1985 and 1986, we have for each incentive system (read state) data on the number and the size of the training agencies, or more simply agents, and on the agents' performances outcomes and awards.³

To motivate the model, we present some basic statistics on the number of agents per state, and on the agents' budgets, awards and performances. The number of agents varies across states. In fact, there are on average 11.9 agents per incentive systems with a standard deviation of 11.0. The average agent's size also varies considerably across incentive systems. Agents manage on average a budget of \$3,084,309 but the standard deviation in average budget across states is \$3,254,630. This variation illustrates the fact that the JTPA funds are allocated to the states by formula on the basis of the relative size of their population that is eligible for training. Those states that have larger eligible

¹For a description of JTPA see Johnston, 1987.

²For a description of the JTPA incentives see Courty and Marschke, 2000.

³The data on the agents' performance outcomes and performance standards used in this study come from the JTPA Annual Survey Report (JASR). This report is compiled annually by the Department of Labor. The award and budget were collected by SRI, International (SRI) and Berkeley Planning Associates to evaluate for the National Council for Employment Policy the efficacy of performance standards in JTPA. See Dickenson, *et al.* (1988) for a description of the data. We thank Carol Romero of the National Commission for Employment Policy for making these data available to us.

population manage more and/or larger training agencies. Agents' budgets within a state can also vary tremendously. The within-state variance in budgets is lower than \$1m in some states and as large as \$10m in others. This, again, is due to the fact that each training agency receives a share of its state's budget that is proportional to its fraction of the state population that is eligible for JTPA training. Most importantly for our study, this variation in the number of agents and in their budgets is exogenous since it depends on the local density of population in need.

As an aside, note that this feature of agent heterogeneity prevails in government organizations where the sizes of the basic managerial entities are largely determined by administrative boundaries. This implies that government organizations typically supervise pools of heterogeneous agents. In fact, this is the case in education (agents are schools), health (hospitals) and many other government service organizations where some experimentation with incentives has been tried (Dixit, 1999).

Next, we describe the performance outcomes. Before presenting some numbers, it may be useful to describe the concept of performance measures and performance standards in the JTPA organization. In fiscal years 1985 and 1986, there were seven performance measures and the U.S. Department of Labor (DOL) required that the States use them all. There were four measures for the adult participants, and three for the youth participants. Table 1 defines the seven performance measures.

Each state in JTPA develops an incentive system based on the DOL-defined measures to reward its pool of training agencies. The states have considerable latitude in the construction of the incentive scheme as long as awards are contingent on the achievement of numerical standards defining minimum acceptable level of performance. For non-cost measures (see Table 1), agents receive awards if their outcomes exceed the corresponding standards. For cost measures, on the other hand, agents receive awards if their outcomes are exceeded by the corresponding standards.

The DOL sets performance standard benchmarks for each performance measure based on the historic performance of other training centers in the system. For the non-cost (cost) measures, the DOL sets the benchmark at the 25th (90th) percentile of the agent performance nationwide for the previous two fiscal years; this means that 75 (90) percent

of agents in the previous two years would have attained the standard. The DOL offers states a procedure for adjusting the each measure's benchmark by the characteristics of the local labor market (e.g., the local unemployment rate) and by characteristics of the agent's enrollee population (e.g., enrollee representation of welfare recipients). The purpose of the adjustment procedure is to level the playing field so that agents are held to standards that are appropriate to their local economic conditions and the kinds of clients served. The states have discretion over the formulation of the standards, but most states during the period under investigation adopted the same DOL formulae to control for outside factors.⁴ Table 2 computes the fraction of agents who have exceeded the performance standard and the average performance in excess of the standard (that is, the actual performance outcome minus the standard) for the seven performance measures. Table 2 shows that while most agents exceed the standard, their excess performances vary considerably.

Finally, we present the award prizes. By mandate, a state's award pot is about seven percent of the training budgets it supervises.⁵ Table 3 presents the mean and standard deviation of the agents' awards, and of their awards per unit of budget. The award per unit of budget varies across agents suggesting that the award funds are not allocated only according to a proportional sharing rule. We also find (not reported here) that the level of awards vary greatly across agents within a state rejecting a fixed sharing rule where the award pots would be distributed equally across agents.

Although the awards vary greatly across agents, there are some important restrictions on the award distribution. First, the awards have to be positive, meaning that the states cannot reduce the agents' budgets following a poor performance. Second, the states cannot spend more than the award budget even if all agents do exceptionally well: the

⁴See Heckman et al. (1997) for a general discussion on the use of performance standards in government organizations.

⁵The JTPA funds are allocated in three sub-funds: 78 percent are set aside for training services, 6 percent are set aside for the incentive system and the remaining 16 percent are set aside for other special services. The award fund as a fraction of total training budget is 7.1 percent ($6/(78+6)$) if one assumes that all award funds are eventually distributed as training budget. The actual figure should typically be lower than 7.1 percent because some of the incentive set aside fund is spent to administrate the incentive funds. In our data, the award as a fraction of budget also varies across states because some agents are missing in some states. The fraction of award to budget will be greater than seven percent, for example, when poorly performing agents are missing.

award has to be fully-funded.⁶

3 The Model

The previous Section showed that budget sizes, performance outcomes, and award prizes varied greatly both within and across incentive systems. One goal of this paper is to investigate whether incentive theory can explain these variations. Our objective in this Section is to provide a framework for structuring and motivating the empirical analysis. In the core of this Section, we restrict to the simple design problem with only two agents. To establish a comparison benchmark, we will also ignore scale effects in budget size. Toward the end of this Section, we show how the main qualitative predictions generalize to multi-agents and non-linear budget effects.

Agent $i \in I = \{1, 2\}$ manages budget b_i with $b_1 \geq b_2$. Agent i has reservation utility $U(b_i) = b_i U$, and exerts effort e_i at cost $b_i c(e_i)$ with c' , c'' , and c''' positive and $c(0) = c'(0) = 0$. The principal values effort $\pi(b_i, e_i) = b_i e_i$ from agent i . Let W denote the award pot for agents b_1 and b_2 .⁷

Budget multiplies all the fundamental parameters of the model in a proportional fashion. The cost and profit functions say that effort is measured in efficiency units. Under no scale effect, effort should be understood as a measure of quality of managerial decision. This framework suggests a simple comparison benchmark corresponding to the efficient (or first-best) levels of effort in the absence of moral hazard problems. The efficient efforts maximize the weighted sum of efforts $b_1 e_1 + b_2 e_2$ subject to the participation constraints $w_i - b_i c(e_i) \geq b_i U$ for $i \in I$ and the budget constraint $W \geq w_1 + w_2$ where w_i is the wage paid to agent $i \in I$. The optimal level of effort is the same for both agents,

$$e^e = c^{-1}(f - U),$$

⁶States may be able to transfer some award fund from one fiscal year to the other although there are some constraints restricting the amount states can transfer. For simplicity, we will focus in the model section on the polar case where the amount they can transfer is zero.

⁷As a side comment, we assumed that the award fund was fixed. This assumption simplifies the analysis and does not really matter for our empirical application since the interest there is not on the optimal award pot (W) but rather on the optimal award function to be defined below. We could also solve for the optimal award pot. This would just add another decision variable without much supplementary insight for our application.

where $f = \frac{W}{b_1+b_2}$ represents the award as a fraction of budget and we assume $f - U > 0$ to guaranty that $e^e > 0$. One should think of effort as an efficiency multiplier in the use of the budget. Both agents supply the same effort because they equally increase the efficiency of their budgets. Agent i 's wage is equal to its relative share of total budget $\frac{b_i}{b_1+b_2}W$.

Next, consider the moral hazard case. In line with the moral hazard paradigm, we assume that the principal cannot directly observe the agents' efforts but observes only an imperfect measure of performance. To simplify, we assume that the performance measure can only take high or low values. Four performance outcomes may occur that we will denote $J = \{hh, hl, lh, ll\}$ where performance outcome hl , for example, is interpreted as agent one performing high and agent two low. Outcome $j \in J$ occurs with probability $p^j(e_1, e_2)$ and agent $i \in I$ then receives w_i^j . To focus on the main issues, we will assume a simple symmetric linear functional form for the joint probabilities. The symmetry and linearity assumptions in addition to the condition that the probability that an agent achieves a given level of performance does not depend on the other agent's effort (e.g. $\frac{d}{de_2}(p^{hh} + p^{hl}) = 0$) imply that $p^{hh}(e_1, e_2) = k^{hh} + \alpha e_1 + \alpha e_2$, $p^{hl}(e_1, e_2) = k^{hl} + \beta e_1 - \alpha e_2$, $p^{lh}(e_1, e_2) = k^{lh} - \alpha e_1 + \beta e_2$, and $p^{ll}(e_1, e_2) = k^{ll} - \beta e_1 - \beta e_2$ with $\alpha, \beta > 0$, and k^j given constants such that $k^{hl} = k^{lh}$ and $p^j \in [0, 1]$ within the relevant effort ranges.

Define agent i 's expected award as,

$$W_i(e_1, e_2) = \sum_{j \in J} p^j(e_1, e_2) w_i^j.$$

To focus on the main issues, we will assume that the agents are risk neutral.⁸ Agent i 's utility under the above award scheme is,

$$U_i(e_i) = W_i(e_1, e_2) - b_i c(e_i).$$

The incentive compatibility constraint for agent i says that she chooses the level of effort that maximizes her utility given the other agent's effort. The first order condition

⁸Under the strong participation constraints to be introduced below this assumption is not very restrictive since the agents are guaranteed their reservation utilities anyway.

to agent i 's maximization problem is,

$$\frac{d}{de_i} W_i(e_1, e_2) = b_i c'(e_i) \quad (ICC_i).$$

The first order condition is sufficient because the agent's maximization problem is convex. The next set of constraints says that the principal guarantees the agents their reservation utility under every performance outcome. Stretching the contract literature's terminology, we will call these constraints the strong participation constraints,

$$w_i^j - b_i c(e_i) \geq b_i U \quad (SPC_i^j),$$

for $j \in J$ and $i \in I$. These participation constraints are stronger than the ones found in the incentive literature, or weak participation constraints, saying that the agents are better-off participating *on average*,⁹

$$U_i(e) \geq b_i U \quad (WPC_i).$$

The final set of constraints is new to this problem and will play an important role in the analysis. These constraints say that the total award payments in any performance outcome cannot exceed the total award pot. We call these constraints the strong budget constraints.

$$W \geq w_1^j + w_2^j \quad (SBC^j),$$

for $j \in J$. The strong budget constraints emerge, for example, when the incentive system has to be fully funded so that the principal cannot transfer award funds from one contract year to the other. They are the mirror image to the principal of what the strong participation constraints are to the agent. The strong budget constraints are stronger than the standard budget constraint found in the incentive literature, or weak budget constraint in this work, saying that the award cannot exceed *on average* the total award pot,

$$W \geq W_1(e_1, e_2) + W_2(e_1, e_2) \quad (WBC).$$

⁹The SPC as modelled here are a strong version of the limited liability constraint found in the literature saying that the agent's utility has to be greater than a fixed constant that could be lower than the agent's reservation utility. SPC occur in practice when the principal needs to overcome the agent's resistance to the introduction of explicit incentives. The principal uses SPC to reassure the agent that she will not lose-out under the new compensation contract (e.g. Lazear, 1999).

In the analysis Section, we will pay special attention to two incentive mechanisms that have received much attention in the contract literature and that are commonly used in practice: piece rate awards and tournaments. An issue of interest will be to investigate if the optimal mechanism can be implemented by these mechanisms. For clarity, we formally define these two mechanisms. A piece rate award mechanism rewards each agent based on her performance outcome alone. Formally, agent b_1 is rewarded according to a piece rate if $w_1^{hh} = w_1^{hl}$ and $w_1^{ll} = w_1^{lh}$. A tournament mechanism ranks the agents and rewards them a prize that depends on their rankings alone. This implies that $w_1^{hl} = w_2^{lh}$ and $w_2^{hl} = w_1^{lh}$.¹⁰

4 Analysis

We analyze the problem gradually. First, we solve the incentive design problem under moral hazard with only the *weak* participation and budget constraints. The novel twist in this analysis is to revisit the standard incentive design problem with heterogeneous agents. Second, we investigate the problem with the *strong* version of these constraints. This is the main contribution of this theoretical section.

Moral Hazard with WBC and WPC Under moral hazard, the efficient outcome can be achieved as long as the *ICCs* and the *WPC* hold at the efficient level of effort. Then, the *WBC* is implied by the *WPCs*. The *ICCs* will hold at the optimal level of effort if the principal can create an award differential between high and low performances large enough to provide the right effort incentives. The principal will be able to bind the *WPC* if it can adjust the average level of performance by punishing the agent under low performance to compensate for the high rewards under high performance.

This will typically be the case as long as the principal has enough degrees of freedom on the 8 outcome dependent awards (w_i^j) to satisfy the 5 constraints ($(ICC_i, WPC_i)_{i \in I}, WBC$). Many mechanisms implement the efficient outcome but the goal of this section is to focus on piece rate and tournament.

¹⁰We assume that when both agents achieve the same outcome, they are randomly ranked. Tournament then implies that the total award given when agents perform the same is equal to the total award when they perform differently, $w_1^{hh} + w_2^{hh} = w_1^{ll} + w_2^{ll} = w_1^{hl} + w_2^{hl}$.

To start, note that the principal cannot implement the efficient outcome under a “pure” tournament. A tournament offers a fixed prize schedule that is independent of the size of contestants. The tournament’s winner then earns the same prize whether it is managing a large or a small budget. When $b_1 > b_2$, tournaments give too much incentive to the small agent relative to the large one. This result is similar to the result in the tournament literature that tournaments may not achieve the efficient outcome when one agent has a comparative cost advantage (Lazear and Rosen, 1981). The solution in these models is to handicap the favorite agent. In our model, a simpler solution consists in a modified tournament structure where the prize schedule is weighted by the sizes of the agents. Define a ‘weighted tournament’ mechanism as a tournament where the winner earns $b_i w^W$ and the loser $b_{-i} w^L$ where w^W and w^L are the prizes per unit of budget.

Proposition 1 *Under WBC and WPC, the efficient outcome can be implemented under a weighted tournament system where awards are proportional to budget sizes.*

A similar analysis applies to piece rate system. Although the principal cannot implement the efficient level of effort with a single piece rate rewarding only high and low performances, she can implement the efficient levels of effort under a weighted piece rate system. Weighted tournament and weighted piece rate belong to a more general class of ‘weighted mechanisms’ that satisfy $\frac{w_1^j}{w_2^j} = \frac{b_1}{b_2}$ for $j = hh, ll$ and $\frac{w_1^{hl}}{w_2^{hl}} = \frac{w_1^{lh}}{w_2^{lh}} = \frac{b_1}{b_2}$. There are many weighted mechanisms that implement the efficient outcome. The intuition is that under a weighted incentive scheme ICC_1 is equivalent to ICC_2 and similarly WPC_1 is equivalent to WPC_2 . Therefore, the principal can achieve the efficient outcome because she has 4 degree of freedom (the four prizes) and must satisfy only two constraints (ICC and WPC). Note, however, that there are some mechanisms that do not satisfy the condition for a ‘weighted mechanism’ and that still implement the efficient outcome.¹¹

Moral Hazard with SBC and SPC Let’s now turn to the design problem with the

¹¹In any mechanism that implement the efficient outcome, the following condition must hold $\frac{\alpha(w_1^{hh} - w_1^{lh}) + \beta(w_1^{hl} - w_1^{ll})}{\alpha(w_2^{hh} - w_2^{lh}) + \beta(w_2^{hl} - w_2^{ll})} = \frac{b_1}{b_2}$. This condition says that prizes have to be weighted but only in an average sense.

strong budget and participation constraints. The incentive design problem is,

$$(ID) \quad \begin{aligned} & \text{Max} \quad \pi_1(e_1) + \pi_2(e_2) \\ & \quad (e_i, w_i^j)_{i \in I}^{j \in J} \\ & \text{s.t.} \quad (ICC_i, SPC_i^j, SBC^j)_{i \in I}^{j \in J} \end{aligned}$$

To start, we consider the relaxed incentive design problem (RID) where we take into account only the inequality $\frac{d}{de_i} W_i(e_1, e_2) \geq b_i c'(e_i)$ from the ICCs. It will be easy to check that the principal can still implement the optimal *RID* profits when the reverse inequalities are imposed.

Lemma 1 *The optimal RID profits can be implemented by an incentive system where $SPC_1^{ll}, SPC_2^{ll}, SPC_2^{hl}, SPC_1^{lh}, SBC^{hh}, SBC^{hl},$ and SBC^{lh} bind (hold as equality) and $SPC_1^{hl}, SPC_2^{lh},$ and SBC^{ll} do not bind.*

To provide effort incentives, the principal tries to create the largest award differential between high and low performances. This has straightforward implications for the states of the world where only one agent performs well. The agent who does not perform gets her reservation utility while the agent who does perform get the rest of the award pot. Similarly, when both agents perform poorly they get only their reservation utilities. Lemma 1 greatly simplifies the incentive design problem. In fact, we can replace, or get rid of, most of the constraints and are left only with $ICC_1, ICC_2, SPC_1^{hh},$ and SPC_2^{hh} . Define the simplified relaxed incentive design problem as,

$$(SRID) \quad \begin{aligned} & \text{Max} \quad \pi_1(e_1) + \pi_2(e_2) \\ & \quad (e_1, e_2, w_2^{hh}) \\ & \text{s.t.} \quad ICC_1, ICC_2, SPC_1^{hh}, SPC_2^{hh} \end{aligned}$$

where the all award prizes but w_2^{hh} have been replaced using Lemma 1. Let λ^c represent the Lagrange multiplier associated with constraint c .

Lemma 2 $\lambda^{SPC_1^{hh}} = 0$ and $\lambda^{ICC_1} > 0$ in the optimal *SRID* contract.

The large agent is the one who is difficult to motivate. The incentive compatibility constraint will always bind for that agent. Similarly, that agent will always receive more than its reservation utility when both agents are performing high. The intuition is simple.

The small agent gets a disproportionately large award when she is the only high performer. Therefore, the small agent is facing stronger incentives than the large one from the way the award pot is distributed when there is only one high performer. This has to be balanced if one wants the two agents to provide the same effort and the only opportunity to over reward the large agent is when both agents perform well. A final result will help interpret the results.

Lemma 3 $\lambda^{SPC_2^{hh}} = 0 \Leftrightarrow e_1 = e_2$ and $\lambda^{SPC_2^{hh}} > 0 \Leftrightarrow e_1 < e_2$.

This Lemma says that the small agent supplies more effort than the large one if she just receives her reservation utility in the state of the world where both agents perform high. The optimal incentive scheme depends on which constraints out of ICC_2 and SPC_2^{hh} bind and this in turn depends on the parameters of the model. Three mutually exhaustive cases may occur. (A formal proof is presented in the Appendix.)

1. Contract (C1), ($\lambda^{ICC_2} > 0, \lambda^{SPC_2^{hh}} = 0$). The solution to *SRID* without SPC_2^{hh} does satisfy SPC_2^{hh} . Then, both agents supply the same effort $e_1 = e_2$. The optimal pair (e, w_2^{hh}) is obtained by solving the agents' first order conditions.
2. Contract (C2), ($\lambda^{ICC_2} > 0, \lambda^{SPC_2^{hh}} > 0$). The small agent supplies more effort than the large one and is paid her reservation utility in the state of the world where both agents perform well $w_2^{hh} = b_2(U + c(e_2))$. The small agent's ICC binds.
3. Contract (C3), ($\lambda^{ICC_2} = 0, \lambda^{SPC_2^{hh}} > 0$). Again, the small agent supplies more effort than the large one and is paid her reservation utility in the state of the world where both agents perform well. The difference now is that the small agent's ICC does not bind. As a consequence the awards w_2^{ll} and w_2^{lh} are not uniquely determined.¹²

Note that the optimal contract is not uniquely determined only in contract (C3) for the small agent and for performance outcomes lh and ll . The intuition for this result is

¹²The optimal *SRID* award scheme violates ICC_2 's reverse inequality. To meet that constraint, it is necessary to lower w_2^{lh} and/or increase w_2^{ll} . It is possible to do so because ICC_2 in *SRID* does not bind so SBC^{lh} and SPC_2^{ll} do not have to bind. Any combination of w_2^{ll} and w_2^{lh} that binds ICC_2 and satisfies SPC_2^{ll} and SBC^{lh} implements the *SRID* profits and satisfy all the *ID* constraints.

simple. The small agent would be facing too powerful incentive if she would receive the entire leftover award pot (after giving the large agent her reservation utility) when she is the only high performer and only her reservation utility when both agents perform low. Under such powerful incentive, the small agent would supply too much effort relative to the large one. Therefore, $\lambda^{ICC_2} = 0$. One solution to lower the small agent's effort is to waste some award funds when the small agent is the only high performer. Another way to go is to increase the small agent's award when both agent perform poorly. The principal is indifferent between these two options.

The Optimal Award Prizes

	C1	C2	C3
w_1^{hh}	$W - w_2^{hh}$ ^(a)	$W - b_2(U + c(e_2))$	$W - b_2(U + c(e_2))$
w_2^{hh}	w_2^{hh}	$b_2(U + c(e_2))$	$b_2(U + c(e_2))$
w_1^{hl}	$W - b_2(U + c(e))$	$W - b_2(U + c(e_2))$	$W - b_2(U + c(e_2))$
w_2^{hl}	$b_2(U + c(e))$	$b_2(U + c(e_2))$	$b_2(U + c(e_2))$
w_1^{lh}	$b_1(U + c(e))$	$b_1(U + c(e_1))$	$b_1(U + c(e_1))$
w_2^{lh}	$W - b_1(U + c(e))$	$W - b_1(U + c(e_1))$	w_2^{lh} ^(b)
w_1^{ll}	$b_1(U + c(e))$	$b_1(U + c(e_1))$	$b_1(U + c(e_1))$
w_2^{ll}	$b_2(U + c(e))$	$b_2(U + c(e_2))$	w_2^{ll}

^a w_2^{hh} solves ICC_1 and ICC_2 for $e_1 = e_2$.

^b Any w_2^{lh} and w_2^{ll} that satisfy SPC_2^{ll} , SBC^{lh} and ICC_2 at the optimal levels of effort (e_1, e_2) .

Table 1 presents the optimal award prizes under the three possible contracts. In contract (C1) when both agents are doing well, the large agent receives a larger award than the smaller one by a factor that overstates their sizes difference ($\frac{w_1^{hh}}{b_1} > \frac{w_2^{hh}}{b_2}$). The intuition is that the small agent is already facing pretty strong incentives because she can be generously rewarded when she is the only high performer. Therefore, the small agent does not need to be rewarded as much as the large one does when both perform well. This result will also typically hold for contracts (C2) and (C3) as long as the small agent does not exert much more effort than the large one.

Table 1 shows that the principal does not always distribute the entire award pot. This will typically occur when performance is low across the board. Burning out some award money is the optimal punishment scheme to provide ex-ante incentives. The rational for this outcome is that the principal cannot carry award funds from one incentive contract to the other. Under contract (C3), the principal may even burn some award fund in the

state of the world where only the small agent performs well.

Another implication of Table 1 is that the optimal incentive scheme cannot be implemented under a (weighted) piece rate system. In fact, under a piece rate system the small agent would receive the same prizes when she is the only high performer and when both agents perform high. In the optimal contract, however, the small agent receives less when both agents perform high than when she is the only high performer ($w_2^{lh} > w_2^{hh}$). Similarly, a (weighted) tournament system cannot be optimal because it would entail to sometimes reward the large agent more than its reservation utility when both agents perform low.

The agents' awards depend not only on their performances but also on the performances of the other agent. The reason for the optimality of group incentive in this model with *SPC* and *SBC* constraints is distinct from the standard reason found in the incentive literature. The traditional reason is that group incentives allow the principal to better insure the agents against performance risk when the measures of performance are stochastically related across agents. This is also known as Holmstrom's (1979) informativeness principle. In this model, agents are risk-neutral and group incentives are optimal even when the performance outcomes are independent across agents. The reason for the optimality of group incentives here comes from the need to cross-subsidize performance rewards in order to increase the award differential in the presence of the strong budget constraint.

In the empirical section, we want to investigate how the optimal contract changes as agents are more heterogeneous and as total budget changes. To investigate this issue theoretically we assume that the budgets are $b_1 = \bar{b} + \Delta_b$ and $b_2 = \bar{b} - \Delta_b$ with $\bar{b} > \Delta_b > 0$. To control for scale effects, we will assume that $W = f\bar{b}$ so that the award pot increases proportionally with budget.

Proposition 2 *There exist $0 < \Delta^* < \Delta^{**} < 1$ such that (C1) is optimal for $\frac{\Delta_b}{\bar{b}} < \Delta^*$, (C2) is optimal for $\Delta^* < \frac{\Delta_b}{\bar{b}} < \Delta^{**}$, and (C3) is optimal for $\frac{\Delta_b}{\bar{b}} > \Delta^{**}$.*

This Proposition together with Lemma 3 implies that $e_1 = e_2$ for $\Delta^* \geq \frac{\Delta_b}{\bar{b}}$ while $e_1 < e_2$ for $\frac{\Delta_b}{\bar{b}} \geq \Delta^*$. When agents are heterogeneous enough, the small agent exerts more

effort than the large one. This will happen if the award system over-rewards the small agent when she is the only high performer so much that this cannot be compensated by under-rewarding her when both agents perform well. The principal will not be able to level incentives across agents when the agents' budgets are too heterogeneous. Note that the proper measure of agent heterogeneity is relative budget difference scaled by mean budget. Budget differences matter more when mean budget is lower. Put differently, the larger the average award pot, the easier it is for the principal to compensate for budget heterogeneity. The final result regards the average level of prizes and the average level of performance.

Proposition 3 *The small agent is more likely to perform high than the large one. The small agent earns more on average than the large one when $\frac{b_1}{b_2} > \frac{p^{hl} + p^{hh}}{p^{lh}}$.*

We conclude with a comment on the welfare implications of the model. The SPC and SBC are source of two kinds of distortions. First, the optimal incentive system does not always allocate effort optimally across agents. When agents are too heterogeneous, the large one exerts too little effort and the small one too much effort. Second, even contract C1 does not achieve the efficient outcome although it does satisfy the condition that both agents supply the same level of effort ($e_1 = e_2$). There are two reasons for that. One reason is that the SBCs force the principal to throw away award funds when both agents perform poorly.¹³ Another reasons is that the agents receive more than their expected reservation utility under SPC. As a consequence, agents exert less effort under SPC and SBC than under WPC and WBC. Note that the inefficiency of having the SBC and the SPC is not driven by one of these constraints alone. In fact, the principal would be better-off with SBC and WPC or with WBC and SPC than with SBC and SPC. Both the limited liability and the fully funded constraints bind.

Extensions The most crucial assumption in the model is the assumption that there are no (dis)economies of scale in budget size. To investigate the role this assumption, we

¹³To the extent that the award money could be used for other activities than agent compensation, the efficiency impact of this distortionary effect could be mitigated and really depends on the value of these other activities. Interestingly, the JTPA incentive system anticipated that potential problem and created a "technical assistance" fund. States are allowed to channel some of the award money to the technical assistance fund to help poorly performing training agencies.

assume that the cost function is not linear in budget ($\frac{\partial^2 C}{\partial^2 b} \neq 0$) and consider how this would change the optimal contracts and the levels of effort. Assuming diseconomies of scale in budget $\frac{\partial^2 C}{\partial^2 b} > 0$ would add a force pushing toward requiring more effort from the small agent relative to the large one while assuming economies of scale in budget $\frac{\partial^2 C}{\partial^2 b} < 0$ would push toward relatively less effort from the small agent. Under moral hazard with SPC and SBC these effects would just add to the incentive effect we identified in the analysis.

It is clear that it would be impossible to identify the incentive effect in a single contract environment without knowing anything about the cost function. This is not true, however, if one has access to a cross section of contracts that cover different pools of agents. To illustrate this point, assume for example, that there are two pairs of agents where the large agent in one pair is the same size as the small agent in the other pair. Then, a simple extension of the model would predict that although these two agents are identical, they should receive different awards when they perform well and their paired agent also do so. In such event, the agent that is paired with a larger agent should receive a smaller award than the one that is paired with a smaller one. All the other predictions of the model can also be identified.

Another important assumption of the model is that there are only two agents. To simplify, consider the case of four agents corresponding to two identical pairs of agent and let's compare this four-agent case (two pairs) with the corresponding two-agent case (one pair). In the four-agent case, the distortion effect identified in the two-agent case will be less pronounced because the principal will have more degree of freedom to smooth the award function across agents. In addition, the performance outcome where all agents perform poorly will occur less frequently implying that the principal will burn out the award pot less frequently. These two forces imply that average performance should increase with the number of agents.

5 Application to the JTPA Incentive System

In this section we test whether states implement the optimal award scheme. The theoretical model establishes how the agents' awards should depend upon their budgets and upon their performances. The theoretical model also makes predictions about how the award distribution and the performance outcomes should vary across states that supervise different agent pools. We test the following predictions of the optimal incentive system:

1. **Award as a function of budget**—An agent that is small relative to the average agent in the state receives disproportionately large awards, given its performance. States should distribute on average less than their entire award pot.
2. **Award as a function of performance**—The agent's award should depend positively on its performances but negatively on the performance of other agents within the same incentive system.
3. **Performance as a function of budget**—Smaller agents should perform better on average than larger ones. States that are more heterogeneous should perform worse.

To test these implications we use data that contain information on performance outcomes on the seven DOL measures, on awards and on budgets. Depending on the prediction we are testing, our unit of observation is either a training agency or a state. Our two data sources were presented in footnote 3. From the SRI data set, we have financial data for approximately 400 of the training agencies in fiscal years 1985 and 1986.

For about 42 states, we have a significant fraction of the agents. The sample we work with represents only about two thirds of the JTPA population of training agencies (recall that there are over 600 training agencies in JTPA) primarily because many training agencies failed to report their awards and/or their budgets. In addition, we have agency performance outcomes and standards for most agencies between 1984 and 1988 from the JASR data set. Broadly speaking, our testing strategy is to examine whether incentive theory predicts how awards are distributed and how agents respond to awards in JTPA.

Award as a function of budget—We begin by testing how the agent’s budget influences its award. In this reduced form approach we focus on the predictions that (a) larger agents should receive larger awards and (b) agents who are relatively small in their states should receive disproportionately larger awards. Model I in Table 4 regresses award on budget. Model I shows that the award rises on average 4 cents for a 1 dollar increase in budget. (The coefficient estimate is statistically significant.) Model II in Table 4 adds to the right hand side of the regression the mean budget in the agent’s state. The mean budget picks up the effect of the agent’s relative size.

Several implications can be drawn from the two regressions in Table 4. First, note that the intercept, which is positive and significant in Model I, is not significantly different from zero in Model II. This says that states do not give fixed prizes independently of size. Second, the coefficient estimate on mean budget is significant and positive, indicating that agents that are large relative to their state peers earn less. These results are consistent with the thrust of the theory.

Award as a function of performance—In testing for budget effects, we concentrate on the determinants of scaled awards or awards per unit of budget. Table 5 explores the implications of the model by examining the effects of performance, and performance relative to the performance of other agents in the state, on the award *as a fraction of the budget*. The regressions in Table 5 include on the right hand side measures of *excess performance*, the agent’s performance outcome minus the corresponding performance standard. The wage and cost measures in the excess performance calculations are denominated in dollars. The employment rate and youth positive termination rate measures are multiplied by 100.¹⁴

In these regressions, the right-hand side contains seven measures of agent excess performance.¹⁵ Recall that the incentive system rewards cost outcomes only when they are exceeded by the cost standard. For the sake of consistency, we compute excess performance for the two cost measures, CE and CEY, as the performance standard minus the

¹⁴For example, the excess adult employment rate measure for an agent who produces a year-end employment rate of 70 percent and faces a standard of 67 percent, is calculated as $70 - 67 = 3$.

¹⁵Each performance measure must receive a positive weight in the determination of the agent’s award.

outcome. That way, if the regression is correctly specified, and each performance outcome matters for the award, we should find that the coefficient estimates on excess performance are positive.

Because we have a two-year panel for each agent, we estimate the relationship using a random effects model, i.e., with separate, agent-specific disturbances. All regressions reported include state dummies to control for state variation in other dimensions of the incentive system that affect award size. We build the model in two steps. We first investigate the role of performance, and then investigate the role of relative performance in the determination of the award size.

Model I contains on the right-hand side only measures of excess performance. The coefficient estimates for the average wage at placement measure, the adult cost measure, and the youth employment measure have the predicted signs and are statistically significant by conventional significance criteria. To understand the impact of performance on the award implied by these point estimates, consider the average agent whose budget is equal to \$3 million (the approximate mean budget in our sample). A \$100 reduction in the cost per placement relative to the cost standard raises the agent's award by approximately \$3,300. A 10 cent increase in the wage at placement relative to the wage standard raises the agent's award by \$14,100. A 10 point increase in the agent's youth placement rate relative to the standard raises the agent's award by \$3,600. These figures correspond to arc award elasticities of .37, .97 and .25, respectively.¹⁶

Model II investigates whether awards are determined by relative performance. On the right hand side, we add to the agent's own excess performance the mean values of excess performance in the agent's state. Negative coefficients on the mean values indicate that agents are paid more when the other agents in their state do worse. Here we are testing the model's predictions that the states construct group incentives. The coefficient estimates on the mean values of excess performance in the average wage at placement and the youth cost measures both have the predicted sign and are significant (the p values

¹⁶Another relevant measure is the budget elasticity to performance. These elasticities are about fourteen times smaller since the award represents only about eight percent of the budget. Although these elasticities may seem small, they are not when compared to similar measures estimated from executive compensation (Jensen and Murphy, 1990).

of the two-tailed tests of significance are .004 and .001, respectively). Consider again the agent with a budget equal to the system average of \$3 million per year. Independent of the agent's absolute average wage at placement (youth cost per placement) outcome, the agent wins approximately an extra \$27,000 (\$8,400) when its wage (youth cost per placement) outcome relative to the state average increases by 10 cents (decrease by \$100).¹⁷

A surprising finding that emerges from Table 5 is that not all performance measures are significant. Related to that result, we also find that the explanatory variables do not explain much of the variation in award per unit of budget. The R^2 for Model I is about .256. As a benchmark, the state dummies alone (this regression is not reported) explain about 13 percent of the total variation in the award per unit of budget. Thus, while the R^2 is low in the model, excess performance accounts for nearly half of the explained variation in the award per unit of budget. The addition of the mean values of excess performance in Model II only modestly raise the R^2 (from 25.6 percent to 27.8 percent).

The statistical insignificance of some coefficients on excess performance and more generally, their limited explanatory power, have three possible causes. First, most award policies are highly nonlinear and complex. The low R^2 may reflect that the linear specification imposed in the regressions does not capture well how performance determines the award. Second, an accurate measure of the relationship between award and performance may be difficult to obtain due to measurement error. Administrative data from JTPA data sources are known to contain considerable error.¹⁸ Third, states may be using award funds to meet political objectives rather than incentive objectives as assumed in our model. For example, states may use award funds to redistribute resources to politically-favored agents, or from one geographical area to the other.

Performance as a function of budget—The model predicts that smaller agents should exert more effort and achieve higher levels of performance. The estimates reported in

¹⁷These findings are consistent with the model but they are also consistent with the hypothesis that the contracts use relative performance to control for common shocks. Our data does not reject the hypothesis that performance is statistically related within states.

¹⁸For example, for fiscal year 1986, the JASR and the SRI data set contain measures of the same performance outcomes and standards. These measures are frequently different, and in non-systematic ways.

Table 6 test this hypothesis. Table 6 presents estimates of the determinants of performance with respect to each of the seven performance measures. Table 6 is divided into 7 panels, a panel for each of the seven performance measures. As in Table 5, the dependent variable is defined as excess performance defined as the performance outcome minus the standard for the non-cost measures and the opposite for the cost measures.

To test whether small agents perform better than large ones, we construct a measure of relative size that is equal to the difference between the agent's budget and the mean budget for its state, normalized by the mean budget.¹⁹ We include the budget variable to control for scale effect in the production of the performance outcome. Having done so, we can be sure that the coefficient on the relative budget measure picks up only the performance effect of the agent's size relative to the size of its peers in the state.

Consider first Model I. In the adult employment rate regression (Panel A), the coefficient on the relative budget measure is negative and significant by conventional criteria (the p value is .09). A negative and significant estimate in this specification is consistent with our hypothesis that because they receive stronger incentives, small agents will generate greater outcomes. We find negative but insignificant coefficients on the relative budget variable for the adult welfare employment rate (Panel B), the adult wage at termination (Panel D), and the youth positive termination rate (Panel G) regressions. The coefficients on relative budget size are positive for both cost measure regressions (Panels C and E) and significant for the youth cost measure (its p value is .04). This later finding is inconsistent with the predictions of the model.

A prediction of the model is that relative size should be more important in states where agents are more heterogeneous. Model II estimates separate coefficients on the relative budget size measure for agents in highly heterogeneous states. We use as a measure of state heterogeneity the standard deviation of budget divided by the total allocation of the state. We divide the standard deviation by the state allocation to capture the idea that the larger the agents in a state, the smaller the distortion caused by a given amount of

¹⁹All regressions include state dummies to control for state variation in the other dimensions of the incentive system (e.g., state-specific modifications to the construction of the performance standard).

spread in budget sizes.²⁰ For Model II, we define three indicator variables: δ^{lt25} is equal to one if the heterogeneity measure of the agent's state falls in the lower 25th percentile of the distribution of state heterogeneity outcomes, and equal to zero otherwise. δ^{2575} is equal to one if the heterogeneity measure falls between the 25th and 75th percentiles, and equal to zero otherwise. δ^{gt75} is equal to one if the heterogeneity measure exceeds the 75th percentile, and equal to zero otherwise. The theoretical model predicts that the relative size should have a more pronounced effect on performance the greater the heterogeneity in the state. Therefore, the coefficient estimates on $\frac{B-\bar{B}}{B} \cdot \delta^{2575}$ and $\frac{B-\bar{B}}{B} \cdot \delta^{gt75}$ in Model II are more likely to be more negative than the coefficient on $\frac{B-\bar{B}}{B}$ alone in Model I.

We find negative and significant coefficient estimates for the variable $\frac{B-\bar{B}}{B} \cdot \delta^{2575}$ for both adult employment measures (the p values are .09 and .08, respectively; Panels A and B).²¹ Coefficient estimates were negative and insignificant for the adult wage measure (Panel D) and the youth positive termination rate measure (Panel F). Again, the coefficients estimates from the cost regressions were positive, contradicting the model. (In the adult cost regression, the coefficient estimate on $\frac{B-\bar{B}}{B} \cdot \delta^{gt75}$ is significant, and in the youth cost regression, the coefficient estimate on $\frac{B-\bar{B}}{B} \cdot \delta^{lt25}$ is significant.) Taken at face value, the evidence that relatively small agents face stronger incentives is mixed.

In Table 7, we test whether states that are more heterogeneous perform worse. For our measure of state performance, we compute a weighted average of excess performance, where the weights are the agents' relative sizes. Our measure of heterogeneity is once again the standard deviation of budget size, normalized by the state's budget allocation. We enter on the right hand side the mean budget in the state, to control for any separate scale effect. We estimate the relationship between a state's size distribution and the weighted performance measures using a panel of between 40 to 50 states for fiscal years

²⁰The model does not clearly specify how one should measure heterogeneity when there are more than two agents. We chose to divide the standard deviation in budget by the sum of budgets rather than by the average budget to capture the idea that a greater number of agents will provide the state more degrees of freedom with which to smooth the award function. In any event, we tried different measure of heterogeneity and they give similar results.

²¹The regression estimates shown in Table 6 suggest that relative size matters more in more heterogeneous states, that is, that the coefficient on $\frac{B-\bar{B}}{B}$ interacted with the heterogeneity measure is negative. We have conducted this test formally. While the point estimate of such a test is more often than not negative, we always reject the hypothesis at conventional levels of significance.

1984 through 1988.

As in Table 6 we estimate 7 separate regressions, one for each performance measure. Model I includes only the budget. Model II contains both the budget and the budget heterogeneity variable, defined as before. Considering Model II, the coefficients on the heterogeneity variable in five of the seven regressions are negative, as predicted. The coefficient is both negative and significant in the youth cost and youth employment regressions (Panels E and G). In the two adult employment rate regressions (Panels A and B) the coefficients are positive, but insignificant. Table 7 therefore presents weak evidence consistent with the model: states with more heterogeneous sets of agents perform worse with respect to the performance measures.

To summarize, the evidence provides some confirmation of the theory's implication for how awards should depend on budgets and performance and how performance should depend on pool composition. We find the following. First, we find that the scale prediction holds: larger agents receive larger awards. We also find that relative size matter: agents that are small relative to their state average receive larger awards. Second, we find that while the relationship is not as strong as we would expect, an agent's award is determined by its performance. This finding implies that a real incentive exists, and that awards are not fully determined by political or equity concerns. Third, we find some evidence that a high-performing agent's award is even higher when the other agents in the state perform poorly although again this evidence is not as widespread as it could be. Thus the award function depends on relative performance in a way that is consistent with the theory. Fourth, we find that for some performance measures, relatively small agents perform better than large ones. This finding is consistent with the major implication of the model: that smaller agents face stronger incentives than larger ones. This evidence is mixed, however. For cost measures, relatively larger agents appear to generate higher outcomes, even after controlling for scale effects. Fifth, we find some evidence that effort distortions are greater in states with greater size disparities among agents. We also find that relatively heterogeneous states perform worse than relatively homogeneous states for some measures of performance.

6 Conclusion

This paper studies the provision of incentives in a large federal job training organization for the disadvantaged. In this organization, each state develops a financially-backed incentive system, subject to the constraint that the individual awards cannot be negative (limited liability constraint) and the sum of the awards cannot exceed a fixed award pot (fully funded constraint). With this pot, states reward a pool of training agencies that typically manage different budgets. The training agencies are evaluated on the basis of their performance relative to a fixed set of performance standards. The states have considerable discretion in the construction of the incentive schemes. Piece rates and tournaments, for example, are allowed.

We show that in the presence of the limited liability and fully funded constraints on the award distribution, the optimal award function will not in general elicit the unconstrained efficient level of effort from the agents. The optimal award scheme ‘over rewards’ small agents relative to large ones. Because small agents receive relatively large awards, they put forth inefficiently high levels of effort. We find strong evidence consistent with the prediction that smaller agencies receive greater rewards and mixed evidence that smaller agents exert more effort. As predicted, we find some evidence that inefficiencies are greater in states that are more heterogeneous. Our evidence suggests that constraints on the award distribution lower the overall effectiveness of performance incentives.

Our analysis suggests that the effectiveness of performance incentive depends on the constraints organizations face. Not all organizations have to distribute a fixed award pot to a pool of agents. In many incentive relationships, there is a surplus to be shared (e.g. peasants and landlords share crops, executives and stockholders share stock market value creation, and firms and sales people share sales margins). It is only when there is nothing to be shared that the principal prefers to set aside an award pot rather than, for example, taking the risk of committing to a subjective award formula that may lead the incentive system to bankruptcy. In that respect, the fixed award pot feature distinguishes the incentive design problem that prevails in government organizations.

From a positive point of view, the analysis suggests that the sorting of agents into

pools is an important step in the design of incentive systems. In the same way that grading on a curve works well only in large classes, the use of fixed award pots works better in large and homogeneous pools. Along the same lines, note that incentives would be more effective if states could transfer some of the award pot from one year to the other, thereby relaxing the fully funded constraint, or if states could punish agents by lowering their budgets when they perform poorly, thereby relaxing the limited liability constraint. Any of these solutions to the design challenge we uncovered, however, introduces practical problems of their own.

References

- [1] Courty, P. and Marschke, G. (July, 2000). The JTPA Incentive System: Program Years 1987-1989. Forthcoming in *Performance Standards in a Government Bureaucracy*, edited by James Heckman, W.E. Upjohn Institute for Employment Research.
- [2] Demski, J., Sappington, D., and Spiller, P. (1988) Incentive Schemes with Multiple Agents and Bankruptcy Constraints. *Journal of Economic Theory*, 44, pages 156–167.
- [3] Dickinson, K., West, R., Kogan, D., Drury, D., Franks, M., Schlichtmann, L., and Vencill, M. (September, 1988). Evaluation of the Effects of JTPA Performance Standards on Clients, Services, and Costs. Manuscript, National Commission for Employment Policy, Research Report No 88-16.
- [4] Dixit, A. (December, 1999). Incentives and Organizations in the Public Sector: An Interpretative Review. Manuscript, Princeton University.
- [5] Heckman, J., Heinrich, C., and Smith, J. (1997). Assessing the Performance of Performance Standards in Public Bureaucracies. *American Economic Review*, 87(2):389–395.
- [6] Holmstrom, B. (1979). Moral Hazard and Observability. *Bell Journal of Economics*, 10 (1) pages 74–91.
- [7] Innes, Y. (1990). Limited Liabilities and Incentive Contracting with Ex-Ante Action Choices. *Journal of Economic Theory*, 52, pages 45–67.
- [8] Jensen, M.C. and Murphy, K.J. (1990). Performance Pay and Top-Management Incentives. *Journal of Political Economy*, 98 (2) pages 225–64.
- [9] Kim, S.K. (1997). Limited Liability and Bonus Contracts. In *Journal of Economics and Management Strategy*, Volume 6, pages 899–913.
- [10] Krishna, V., and Morgan, J. (1988). The Winner-Take-All Principle in Small Tournaments. In *Advance in Applied Microeconomics*, Volume 7, pages 61–74.

- [11] Lazear, Edward P. (July 1996). Performance Pay and Productivity. NBER working paper 5672.
- [12] Lazear, Edward P., and Rosen, Sherwin (1981). Rank-Order Tournaments as Optimum Labor Contracts. *Journal of Political Economy*, pages 841–64.
- [13] Moldovanu, B., and Sela, A. (1999). The Optimal Allocation of Prizes in Contests. SonderForschungsBereich 504 Working Paper 99-75, Mannheim University.
- [14] Prendergast, C. (March, 1999). The provision of incentives in firms. *Journal of Economic Literature*, 37(1).
- [15] Sappington, D. (1983). Limited Liability Contracts between Principal and Agent. *Journal of Economic Theory*, 29, pages 1–21.

Proofs

Proof of Proposition 1: First I show by contradiction that pure tournament system cannot implement the efficient level of effort when $b_1 > b_2$. Assume the opposite. Under the efficient level of effort $e_1 = e_2$ while under tournament w^W and w^L the agent's ICC say,

$$\begin{cases} \frac{1}{2}(p_1^{hh} + p_1^{ll})(w^W + w^L) + p_1^{hl}w^W + p_1^{lh}w^L = b_1c'(e_1) \\ \frac{1}{2}(p_2^{hh} + p_2^{ll})(w^W + w^L) + p_2^{hl}w^W + p_2^{lh}w^L = b_2c'(e_2). \end{cases}$$

These condition imply $e_2 > e_1$. A contradiction. QED

Next, define the per-unit of budget tournament prizes,

$$\begin{cases} \frac{1}{2}(p^{hh} + p^{ll})(w^W + w^L) + p^{hl}w^W + p^{lh}w^L = U - c(e^{fb}) \\ \frac{1}{2}(p_1^{hh} + p_1^{ll})(w^W + w^L) + p_1^{hl}w^W + p_1^{lh}w^L = c'(e^{fb}) \end{cases}$$

A solution to this system always exists. The tournament prizes where the winner earns $b_i w^W$ and the loser $b_{-i} w^L$ satisfy the ICCs and WPCs at the efficient level of effort. The WBC is implied by the WPCs.

Proof of Lemma 1: The proof goes by contradiction. Assume for example that SPC_1^{ll} does not bind, i.e., $w_1^{ll} > b_1(U + c(e_1))$. Consider a new contract where w_1^{ll} is decreased such that SPC_1^{ll} binds. SBC^{ll} and ICC_1 still hold while all the other constraints are unchanged. A contradiction. The same reasoning applies to show that there is an optimal contract where SPC_2^{ll} , SPC_2^{hl} , and SPC_1^{lh} bind. Next, we show that SBC^{hh} also binds. Assume it does not. Consider a new contract where w_1^{hh} is increased such that SBC^{hh} binds. SPC_1^{hh} and ICC_1 still hold while all other constraints are unchanged. A contradiction. A similar argument shows that SBC^{hl} and SBC^{lh} also bind. Next, we show that SPC_1^{hl} does not bind. Assume it does, that is, $w_1^{hl} = b_1(U + c(e_1))$. Because SPC_2^{hl} and SBC^{hl} bind $b_1(U + c(e_1)) + b_2(U + c(e_2)) = W$. This implies that $w_i^j = b_i(U + c(e_i))$ for $i \in I$ and $j \in J$. ICC_i imply that $e_i = 0$ for $i \in I$. But $k - U >$ guaranties that there is a solution with positive efforts. A contradiction. The same reasoning shows that SPC_2^{lh} does not bind. Finally, we show that SBC^{ll} does not bind. Assume it does bind, that is, $w_1^{ll} + w_2^{ll} = W$. Since SPC_1^{ll} and SPC_2^{ll} bind, we have $w_i^j = b_i(U + c(e_i))$ for $i \in I$ and $j \in J$. Again a contradiction. QED

Proof of Lemma 2: We focus on solutions with positive efforts for both agents. The RSID problem is,

$$\begin{aligned} & \text{Max } \pi_1(e_1) + \pi_2(e_2) \\ & \quad (e_i, w_2^{hh})_{i \in I} \\ (ICC_1) \quad & p_1^{hh}(W - w_2^{hh}) + p_1^{hl}(W - b_2(U + c(e_2))) + (p_1^{lh} + p_1^{ll})b_1(U + c(e_1)) \geq b_1c'(e_1) \\ (ICC_2) \quad & p_2^{hh}w_2^{hh} + p_2^{lh}(W - b_1(U + c(e_1))) + (p_2^{hl} + p_2^{ll})b_2(U + c(e_2)) \geq b_2c'(e_2) \\ (SPC_1^{hh}) \quad & W - w_2^{hh} \geq b_1(c(e_1) + U) \\ (SPC_2^{hh}) \quad & w_2^{hh} \geq b_2(c(e_2) + U) \end{aligned}$$

The first order condition to RSID are

$$\begin{aligned} FOC_{e_1} & 1 + \lambda^{ICC_1}(p_1^{lh} + p_1^{ll})c'(e_1) - \lambda^{ICC_2}p_2^{lh}c'(e_1) - \lambda^{ICC_1}c''(e_1) - \lambda^{SPC_1^{hh}}c'(e_1) = 0 \\ FOC_{e_2} & 1 + \lambda^{ICC_2}(p_2^{hl} + p_2^{ll})c'(e_2) - \lambda^{ICC_1}p_1^{hl}c'(e_2) - \lambda^{ICC_2}c''(e_2) - \lambda^{SPC_2^{hh}}c'(e_2) = 0 \\ FOC_{w_2^{hh}} & -\lambda^{ICC_1}p_1^{hh} + \lambda^{ICC_2}p_2^{hh} - \lambda^{SPC_1^{hh}} + \lambda^{SPC_2^{hh}} = 0 \end{aligned}$$

Consider first the case $\lambda^{ICC_2} > 0$. We show by contradiction that $\lambda^{SPC_1^{hh}} > 0$ and $\lambda^{SPC_2^{hh}} = 0$ is impossible. Assume this is true. $\lambda^{SPC_1^{hh}} > 0$ implies $w_2^{hh} = W - b_1(c(e_1) + U)$. ICC_2 says that,

$$c'(e_2) = \frac{p_2^{hh} + p_2^{lh}}{b_2}(W - b_1(U + c(e_1)) - b_2(U + c(e_2)))$$

But $\frac{p_2^{hh} + p_2^{lh}}{b_2}(W - b_1(U + c(e_1)) - b_2(U + c(e_2))) > \frac{p_1^{hl}}{b_1}(W - b_1(U + c(e_1)) - b_2(U + c(e_2))) > c'(e_1)$ where the last inequality holds by ICC_1 . Therefore, $c'(e_2) > c'(e_1)$ and $e_2 > e_1$. Next, $\lambda^{SPC_1^{hh}} = -\lambda^{ICC_1}p_1^{hh} + \lambda^{ICC_2}p_2^{hh} > 0$ and the symmetry property saying that $p_1^{hh} = p_2^{hh}$ implies that $\lambda^{ICC_2} > \lambda^{ICC_1}$. Replace $\lambda^{SPC_1^{hh}}$ in FOC_{e_1} and subtract FOC_{e_1} and FOC_{e_2} gives after using the symmetry properties for the marginal probabilities,

$$\lambda^{ICC_1}(p_1^{hl}(c'(e_2) - c'(e_1)) - c''(e_1)) = \lambda^{ICC_2}((p_1^{hh} + p_1^{hl})(c'(e_1) - c'(e_2)) - c''(e_2)).$$

Since $(p_1^{hh} + p_1^{hl})(c'(e_1) - c'(e_2)) - c''(e_2) < 0$, $\lambda^{ICC_2} > \lambda^{ICC_1}$ implies after simplifications,

$$(2p_1^{hl} + p_1^{hh})c'(e_2) + c''(e_2) < (2p_1^{hl} + p_1^{hh})c'(e_1) + c''(e_1).$$

The above inequality implies $e_1 > e_2$. A Contradiction.

Next, we show by contradiction that $\lambda^{SPC_1^{hh}} > 0$ and $\lambda^{SPC_2^{hh}} > 0$ is impossible. This would imply that $W = (b_1 + b_2)U + b_1c(e_1) + b_2c(e_2)$. Then, $w_i^j = b_i(U + c(e_i))$ for $i \in I$ and $j \in J$ and $e_i = 0$. A contradiction.

Finally, we turn to the case $\lambda^{ICC_2} = 0$. Assume $\lambda^{SPC_1^{hh}} > 0$. $FOC_{w_2^{hh}}$ implies $\lambda^{SPC_2^{hh}} > 0$. But $\lambda^{SPC_1^{hh}} > 0$ and $\lambda^{SPC_2^{hh}} > 0$ imply $w_i^j = b_i(U + c(e_i))$ for $i \in I$ and $j \in J$ and $e_i = 0$. A contradiction.

This establishes the Lemma's first claim, $\lambda^{SPC_1^{hh}} = 0$. To establish the Lemma's second claim, plug $\lambda^{SPC_1^{hh}} = 0$ in $FOC_{w_2^{hh}}$ and assume $\lambda^{ICC_1} = 0$,

$$\lambda^{ICC_2}p_2^{hh} + \lambda^{SPC_2^{hh}} = 0.$$

A contradiction since $p_2^{hh} > 0$. Therefore, $\lambda^{ICC_1} > 0$. QED

Proof of Lemma 3: This first part of this Lemma says that the two agents supply the same effort when SPC_2^{hh} does not bind. $\lambda^{SPC_2^{hh}} = 0$ imply $\lambda^{ICC_1} = \lambda^{ICC_2}$. Taking the difference in FOC_{e_1} and FOC_{e_2} gives

$$(p_1^{lh} + p_1^{ll} - p_2^{lh})c'(e_1) - c''(e_1) = (p_2^{hl} + p_2^{ll} - p_1^{hl})c'(e_2) - c''(e_2).$$

After simplification, the above equation implies that $e_1 = e_2$. The second claim in the Lemma naturally follows. QED

Derivation of the Optimal Contract Consider first the case where $\lambda^{ICC_2} > 0$ and $\lambda^{SPC_2^{hh}} = 0$. Then, the optimal level of effort and w_2^{hh} are given by ICC_1 and ICC_2 ,

$$(C1) \begin{cases} p_1^{hh}(W - w_2^{hh}) + (p_1^{lh} + p_1^{ll})b_1(U + c(e)) + p_1^{hl}(W - b_2(U + c(e))) = b_1c'(e) \\ p_2^{hh}w_2^{hh} + (p_2^{hl} + p_2^{ll})b_2(U + c(e)) + p_2^{lh}(W - b_1(U + c(e))) = b_2c'(e) \end{cases}$$

This is the solution to the optimal design problem if the optimal wage satisfy SPC_2^{hh} , that is, $w_2^{hh} - b_2c(e) \geq b_2$. Next we show that $\frac{w_1^{hh}}{b_1} > \frac{w_2^{hh}}{b_2}$. Using ICC_1 and ICC_2 ,

$$\begin{aligned} \frac{w_2^{hh}}{b_2} &= \frac{c'(e) - (p_2^{hl} + p_2^{ll})(U + c(e)) - p_2^{lh}\frac{W - b_1(U + c(e))}{b_2}}{p_2^{hh}} > \\ &= \frac{c'(e) - (p_1^{lh} + p_2^{ll})(U + c(e)) - p_1^{hl}\frac{W - b_2(U + c(e))}{b_1}}{p_1^{hh}} = \frac{w_1^{hh}}{b_1}. \end{aligned}$$

Consider next the case where $\lambda^{ICC_2} > 0$ and $\lambda^{SPC_2^{hh}} > 0$. $\lambda^{SPC_2^{hh}} > 0$ implies $w_2^{hh} = W - b_2(U + c(e_2))$. The optimal levels of effort are given by solving for the agents' first order conditions.

$$(C2) \begin{cases} (p_1^{hh} + p_1^{hl})(W - b_2(U + c(e_2))) + (p_1^{lh} + p_1^{ll})b_1(U + c(e_1)) = b_1c'(e_1) \\ (p_2^{hh} + p_2^{hl} + p_2^{ll})b_2(U + c(e_2)) + p_2^{lh}(W - b_1(U + c(e_1))) = b_2c'(e_2) \end{cases}$$

The first order condition to the design problem are equivalent to $c''(e_1) \geq (p_1^{hh} + p_1^{hl})(c'(e_2) - c'(e_1))$ and $e_2 \geq e_1$.

The final case is $\lambda^{ICC_2} = 0$ and $\lambda^{SPC_2^{hh}} > 0$. After replacement, one can show that the optimal levels of effort are given by,

$$(C3) \begin{cases} (p_1^{hh} + p_1^{hl})(W - b_2(U + c(e_2))) + (p_1^{lh} + p_1^{ll})b_1(U + c(e_1)) = b_1c'(e_1) \\ c''(e_1) = (p_1^{hh} + p_1^{hl})(c'(e_2) - c'(e_1)) \end{cases}$$

This contract does not satisfy ICC_2 's reverse inequality in ID. However, the optimal profits can be implemented under ID by increasing w_2^{ll} and/or decreasing w_2^{lh} by the correct amounts so that ICC_2 holds at the optimal (C3) levels of efforts.

Finally, from $FOC_{w_2^{hh}}$ we have $\lambda^{ICC_2} = 0 \Rightarrow \lambda^{SPC_2^{hh}} > 0$ implying that the case $\lambda^{ICC_2} = \lambda^{SPC_2^{hh}} = 0$ is impossible. The three contracts $C1$, $C2$ and $C3$ are exhaustive and mutually exclusive.

Proof of Proposition 2: Define e^* as the level of effort that solves the agents' ICC s in (C1). Adding the two ICC s gives,

$$kp_1^{hh} + 2(p_1^{lh} + p_1^{ll})(U + c(e^*)) + 2p_1^{hl}(k - (U + c(e^*))) = 2c'(e^*).$$

The above equation shows that e^* does not depend on $\frac{\Delta}{b}$. Define Δ^* such that the agents' ICC s hold at e^* when $w_2^{hh} = \bar{b}(1 - \Delta^*)(U + c(e^*))$,

$$\Delta^* = \frac{p_1^{hh}}{2c'(e^*)}(f - 2(U + c(e^*))).$$

For $\frac{\Delta b}{b} = \Delta^*$, (C1) is obviously the solution to the design problem. Define $G(\Delta b/\bar{b}) = \frac{w_2^{hh}}{b} - (1 - \frac{\Delta b}{b})(U + c(e))$. Using ICC_1 one can show that $\frac{dG}{d(\Delta b/b)} = -\frac{c'(e)}{p_1^{hh}} < 0$. Therefore, the optimal w_2^{hh} in C^1 actually satisfies SPC_2^{hh} when $\frac{\Delta b}{b} < \Delta^*$ and does not satisfy SPC_2^{hh} when $\frac{\Delta b}{b} > \Delta^*$.

Next, define $(e_1^*, e_2^*, \Delta^{**})$ such that,

$$\begin{cases} (p_1^{hh} + p_1^{hl})(W - (\bar{b} - \Delta^{**})(U + c(e_2^*)) - (\bar{b} + \Delta^{**})(U + c(e_1^*))) = (\bar{b} + \Delta^{**})c'(e_1^*) \\ (p_2^{lh}(W - (\bar{b} - \Delta^{**})(U + c(e_2^*)) - (\bar{b} + \Delta^{**})(U + c(e_1^*))) = (\bar{b} - \Delta^{**})c'(e_2^*) \\ c''(e_1^*) = (p_1^{hh} + p_1^{hl})(c'(e_2^*) - c'(e_1^*)) \end{cases}$$

For $\frac{\Delta b}{b} = \Delta^{**}$, (e_1^*, e_2^*) solves both C2 and C3. Next, define $H(e_1, e_2) = c''(e_1) - (p_1^{hh} + p_1^{hl})(c'(e_2) - c'(e_1))$. Using the ICCs, it is possible to show that $\frac{dH}{d(\Delta b/b)} < 0$. For $\frac{\Delta b}{b} < \Delta^{**}$, $H > 0$ measured at the optimal C2 level of efforts and C2 is the optimal contract. For $\frac{\Delta b}{b} > \Delta^{**}$, $H < 0$ measured at the optimal C2 level of efforts and C3 is the optimal contract.

Finally, we show that $0 < \Delta^* < \Delta^{**} < 1$. The first inequality holds because $f - 2(U + c(e^*)) > 0$. The second inequality holds because C1 and C3 are mutually exclusive. The third inequality follows after simplifying the ICCs defining (e_1^*, e_2^*) ,

$$\frac{p_1^{hh} + p_1^{hl}}{p_2^{lh}}(1 - \Delta^{**})c'(e_2^*) = (1 + \Delta^{**})c'(e_1^*)$$

implying,

$$\Delta^{**} = \frac{\frac{p_1^{hh} + p_1^{hl}}{p_2^{lh}}c'(e_2^*) - c'(e_1^*)}{\frac{p_1^{hh} + p_1^{hl}}{p_2^{lh}}c'(e_2^*) - c'(e_1^*)} < 1. QED$$

Proof of Proposition 3: The small agent is more likely to perform high than the large agent if

$$p^{hh} + p^{lh} > p^{hh} + p^{hl}.$$

This is equivalent to $(e_2 - e_1)(\alpha + \beta) > 0$ which is always true.

The small agent earns more on average than the large agent if,

$$\frac{W_1}{b_1} < \frac{W_2}{b_2}.$$

After reordering terms, this is equivalent to,

$$p^{lh}b_1(W - (b_1 + b_2)(U + c(e_1))) - (p^{hl} + p^{hh})b_2(W - (b_1 + b_2)(U + c(e_2))) + p^{ll}b_1b_2(c(e_2) - c(e_1)) \geq 0.$$

$e_2 \geq e_1$ implies that the above inequality always holds when $\frac{b_1}{b_2} > \frac{p^{hl} + p^{hh}}{p^{lh}}$. QED