Dominance of Contributions Monitoring in Teams*

Parimal Kanti Bag† Peng Wang‡

February 14, 2015

Abstract

In team problems it has been argued that there is no loss to the principal from monitoring team output compared to monitoring of individual contributions (McAfee and McMillan, 1991). Optimal output monitoring however requires up front payment from the agents to the principal. By introducing limited liability (LL) on the part of agents that rules out positive monetary transfers to the principal, it is shown that McAfee-McMillan's equivalence result breaks down: principal strictly benefits by monitoring individual contributions.

While superiority of contributions monitoring could be seen as similar to Khalil and Lawarrée's (1994) result on strict dominance of input monitoring in single agent setting, contributions vs. output monitoring is rather a different comparison from that of input vs. output monitoring. In the latter, LL plays no decisive role.

***JEL* Classification:** D82; D86; J33; J41; M52. **Key Words:** Principal, agents, teams, joint projects, adverse selection, moral hazard, limited liability, monitoring, punishing contract.

*Acknowledgement: We especially thank Satoru Takahashi for several very useful suggestions. Any remaining mistakes are ours.

†Department of Economics, National University of Singapore, Faculty of Arts and Social Sciences, AS2 Level 6, 1 Arts Link, Singapore 117570; E-mail: ecsbpk@nus.edu.sg

‡Department of Economics, National University of Singapore, Faculty of Arts and Social Sciences, AS2 Level 6, 1 Arts Link, Singapore 117570; E-mail: ecswp@nus.edu.sg

1 Introduction

McAfee and McMillan (1991) studied a team monitoring problem where the principal can incentivize either by joint output without observing the team members' individual contributions, or by giving rewards based on individual contributions. The team members' (or agents') abilities are private information. The authors show that under appropriate conditions a compensation scheme linear in the team's aggregate output is optimal. That is, the disaggregated information on individual contributions is of no extra value to the principal: the two types of information are equivalent. This is a very intriguing result – with the disaggregated information the principal is expected to monitor more directly the individual agents in a team environment and incentivize them better. Our focus will be to understand this puzzle better and contribute to the broad debate of *input/contributions vs. output monitoring*.

After McAfee and McMillan's paper, Khalil and Lawarrée (1994) made an apparently contrasting observation in a single agent setting with adverse selection and deterministic production. When the principal is the residual claimant (same as in McAfee and McMillan), Khalil and Lawarrée argue that input monitoring (where an agent's effort can be directly contracted upon) is better than output monitoring (with no additional information on agent effort). *Input monitoring* in Khalil and Lawarrée is superficially similar to the monitoring of *individual contributions* in McAfee and McMillan. This suggests that by transiting from the single agent to a team setting, the balance of optimal monitoring shifts towards parity. We will argue that this view is incomplete and to some extent misleading. Rather, there are important differences between Khalil and Lawarrée's concept of input monitoring and McAfee and McMillan's idea of monitoring of individual contributions. And yet, by relaxing one restrictive assumption of McAfee and McMillan we will obtain a similar dominance result as in the single agent formulation of Khalil and Lawarrée.

Specifically, we ask what aspect of McAfee and McMillan's optimal output monitoring mechanism could be critical to their equivalence result, and how plausible that might be. We identify one such feature that plays a decisive role – the principal could ask the agents to pay upfront strictly positive amount of money which they stand to lose if the team output turns out to be "unsatisfactory." That is, effectively, the principal is asking the agents to pay a fee to be able to participate in the team activity. We will depart from this important assumption by requiring that the agents are subjected to limited liability of *not having to make any payment to the principal in any eventuality*. Any positive transfer can only be one-sided – from the principal to the agents and not the other way around. In most team based work arrangements, it is implausible for the main initiator of a project, the principal, to ask its members to contribute to the project and yet to post at the start a bond that they might forfeit if team performances do not go according to plans.[1] Our agents might be financially constrained to make such arrangements feasible. While imposing this limited liability restriction, we still retain McAfee and McMillan's assumption that the agents are risk neutral.[2] With this one modification, we will show that

[1]Exceptions could be law firms or a group of medical doctors in private practices where junior partners may have to pledge compensation at the start in case the firm (or the group) does not perform well. Simon Grant suggested this example.

[2]Vander Veen (1995) had claimed that McAfee-McMillan's equivalence result should break down if the agents

the principal should strictly prefer monitoring individual contributions over output monitoring. Importantly, our result holds whether the production technology involves complementarity or substitution of the agents' contributions, whereas McAfee-McMillan's equivalence result was derived under the complementary production technology assumption.

Our argument for the dominance of contributions monitoring will proceed as follows. We first propose a feasible contributions monitoring contract which could generate the same expected payoff for the agents and the principal as the output monitoring contract does. This contributions monitoring contract is of the take-it-or-leave-it form, with the principal setting a target individual contribution for each declared type of an agent (and a reported type profile of the other agents) and giving him an agreed performance reward only if the set target has been met. This disciplines agents' misbehavior in type reporting as their contribution choices following non-truthful reporting are restricted. We then show that under the optimal output monitoring contract the principal needs to leave positive rent to all types of at least one agent, while under the optimal contributions monitoring contract the principal is able to extract the rent from the lowest type of every agent. Since the optimal output monitoring contract can be replicated by contributions monitoring contract with equivalent expected payoff for the principal as well as all types of all agents, and we know that the replicated contract is not optimal because the lowest type of at least one agent enjoys positive rent, this suggests the superiority of contributions monitoring contract.

A more direct intuition can be given as follows. When limited liability condition is imposed, the principal cannot punish the agents sufficiently for failing to meet their target contributions, which makes inducing any intended contributions more costly. This problem is exacerbated when the agents' contributions cannot be directly observed so that the principal has to rely on output monitoring. In this case, if the output level turns out to be below expectation, the principal is not able to identify who are the main delinquents. In other words, limited liability and moral hazard, together, make it harder for the principal to induce contributions. In the contributions monitoring contract on the other hand, once the principal specifies a type contingent target performance for each agent, any deviation in individual contributions is always detected by the principal. Thus, less information rents are needed under the contributions monitoring contract.

Earlier in one of the first generation analysis of team moral hazard problems, Holmström (1982) had noted that "if there is uncertainty in production and if agents are risk averse or have limited endowments, monitoring becomes an important instrument in remedying moral hazard." In McAfee and McMillan's work, agents are risk neutral and can be asked to make upfront payment, whereas the production technology in Khalil and Lawarrée is deterministic. Modifying McAfee and McMillan's model, we verify Holmström's claim about the relevance of limited endowments by showing that the *type of monitoring* matters.[3] Our analysis should also clarify that (i) limited liability plays no special role in single agent setting for the poor performance

are risk averse. While there is good intuition why incentivizing agents based on individual contributions should be better under risk aversion, Vander Veen's proof does not seem very convincing. See our discussion in the Appendix.

[3]Vander Veen (1995) could be seen as confirming the role of risk aversion, but as pointed out above his proof is doubtful.

of output monitoring, thus team environment lends limited liability an added significance;[4] (ii) monitoring individual contributions in a team, as in McAfee and McMillan, does not require as close a monitoring as that of effort in Khalil and Lawarrée, yet this information is enough to dominate information on collective performance, i.e., output.[5,6]

Some authors studying principal-agent problems under limited liability consider either only the adverse selection problem or only the moral hazard problem. Innes (1990) examined a principal-agent model of financial contracting with moral hazard and limited liability. Sappington (1983) and Lewis and Sappington (2000) derived optimal contracts under adverse selection and limited liability. Lawarrée and Audenrode (1996) highlighted the importance of limited liability in an adverse selection model. They showed that when both the principal and the agent are risk neutral, the optimal contract is the same regardless of whether the output can be perfectly observed or not if the agent has unlimited liability, but the optimal contract changes significantly under imperfect information if the agent's liability becomes limited.[7] Ollier and Thomas (2013) analyze the properties of optimal output monitoring contract with all three features (adverse selection, moral hazard and limited liability).

The rest of the paper proceeds as follows. A single agent example in section 2 illustrates how contribution monitoring might dominate. The formal team model with two types of monitoring contracts is analyzed in section 3. The results comparing the monitoring contracts are presented in section 4. The proofs appear in an Appendix. Detailed proof of Lemma 1 is included in a Supplementary file.

2 Single Agent Monitoring

We start with an example involving a single agent, stochastic production, adverse selection (and limited liability) in order to prepare ourselves for the subsequent analysis of contributions vs. output monitoring. Also, we relate back to Khalil and Lawarrée comparison between input (effort) monitoring and output (contribution) monitoring and present a clear relationship among different monitoring instruments.

A risk neutral principal wants to allocate a task to a risk neutral agent. The agent's marginal productivity z is either high (θ_H) or low (θ_L) with equal probability, and $0 < \theta_L < \theta_H < 1$. Only the agent knows his true (productivity) type. With effort $e \in [0, 1]$, the high type agent's cost is $c_H(e) = \frac{e^2}{2}$ whereas for low type the cost is $c_L(e) = e^2$. Let y_j represent type θ_j's ($j = L, H$)

[4] In Khalil and Lawarrée's deterministic production formulation, limited liability plays no role for the dominance of input monitoring because their optimal mechanisms would satisfy limited liability. In section 2 examples involving a single agent and stochastic production, we establish a more extensive dominance result both with and without limited liability.

[5] On casual thinking it is not clear whether individual contributions are easier to disentangle from aggregate output than monitoring individual effort levels; contribution is the product of one's effort and ability, so even if effort is known determining one's contribution requires the knowledge of ability. But with direct mechanism, truthful revelation of types becomes easier (or less costly) if incentives can be written in terms of effort rather than contribution.

[6] Thus, our analysis of the team problem with stochastic production, along with an intermediate study of a single agent problem, goes beyond Khalil and Lawarrée's single agent model with deterministic production.

[7] The authors do not compare the performance of alternative monitoring mechanisms, as we do, to see the importance of limited liability.

contribution. We assume $y_j = \theta_j e$. The output x, which can be observed by the principal, is stochastic and can be either 0 or 1 with $\Pr(x = 1|y_j) = y_j$. The principal's utility is $U(x) = x$.

The time line is as follows. After the agent learns his type he chooses one of two contracts offered by the principal, one for each type, or decides not to participate. On participation he chooses an effort level, the output is realized, and payment is given according to the contract.

■ **Effort monitoring**. The payment depends on the effort level e as well as the agent's reported type which is denoted by $\hat{z}$, i.e., $\tilde{p}(e, \hat{z})$. Since there are only two types, the equilibrium effort levels are $\tilde{e}_H$ for high type agent and $\tilde{e}_L$ for low type agent, and we can denote the corresponding payments to be $t_H \equiv \tilde{p}(\tilde{e}_H, \theta_H)$ and $t_L \equiv \tilde{p}(\tilde{e}_L, \theta_L)$.[8] Therefore, the principal's objective is

$$\max_{t_H, t_L, \tilde{e}_H, \tilde{e}_L} \frac{1}{2}(\theta_H \tilde{e}_H - t_H) + \frac{1}{2}(\theta_L \tilde{e}_L - t_L)$$

subject to the following incentive compatibility (IC) and participation (PC) constraints:

$$t_H - \frac{\tilde{e}_H^2}{2} \geq t_L - \frac{\tilde{e}_L^2}{2}, \quad t_L - \tilde{e}_L^2 \geq t_H - \tilde{e}_H^2, \quad t_H - \frac{\tilde{e}_H^2}{2} \geq 0, \quad t_L - \tilde{e}_L^2 \geq 0.$$

Solving the above problem yields the optimal payments $t_H = \frac{1}{2}\theta_H^2 + \frac{1}{18}\theta_L^2$ and $t_L = \frac{1}{9}\theta_L^2$, and the efforts are $\tilde{e}_H = \theta_H$ and $\tilde{e}_L = \frac{1}{3}\theta_L$. Therefore, the principal's profit is $E[\tilde{\phi}] = \frac{1}{4}\theta_H^2 + \frac{1}{12}\theta_L^2$, and the agents' payoffs are $E[\tilde{\pi}_H] = \frac{1}{18}\theta_L^2$ and $E[\tilde{\pi}_L] = 0$, where $\tilde{\phi}$ and $\tilde{\pi}_j$ are the principal's ex-post profit and the agent's ex-post payoff respectively. Later on we will use similar notations for payments, profits and payoffs.

■ **Contribution monitoring**. Now, the principal can monitor only the agent's contribution. Thus, the payment depends on the contribution level as well as the agent's reported type, i.e., $\bar{p}(y_j, \hat{z})$. Since there are only two types, the equilibrium contribution levels are $\bar{y}_H = \theta_H \bar{e}_H$ for high type agent and $\bar{y}_L = \theta_L \bar{e}_L$ for low type agent where $\bar{e}_j$ is equilibrium effort of type θ_j agent, and we can similarly denote the corresponding payments to be $w_H \equiv \bar{p}(\bar{y}_H, \theta_H)$ and $w_L \equiv \bar{p}(\bar{y}_L, \theta_L)$. Therefore, the principal's objective is

$$\max_{w_H, w_L, \bar{e}_H, \bar{e}_L} \frac{1}{2}(\theta_H \bar{e}_H - w_H) + \frac{1}{2}(\theta_L \bar{e}_L - w_L)$$

subject to the following constraints:

$$w_H - \frac{\bar{e}_H^2}{2} \geq w_L - \frac{\theta_L^2 \bar{e}_L^2}{2\theta_H^2}, \quad w_L - \bar{e}_L^2 \geq w_H - \frac{\theta_H^2 \bar{e}_H^2}{\theta_L^2}, \quad w_H - \frac{\bar{e}_H^2}{2} \geq 0, \quad w_L - \bar{e}_L^2 \geq 0.$$

Here, when the high type agent mimics as low type, he needs to contribute y_L. Thus, his deviation effort is effectively $\frac{\bar{e}_L \theta_L}{\theta_H}$. Similar result can be derived for the low type agent. Solving the above problem yields the optimal payments $w_H = \frac{1}{2}\theta_H^2 + \frac{2-(\frac{\theta_L}{\theta_H})^2}{2[4-(\frac{\theta_L}{\theta_H})^2]^2}\theta_L^2$ and $w_L = \frac{1}{[4-(\frac{\theta_L}{\theta_H})^2]^2}\theta_L^2$, and the efforts are $\bar{e}_H = \theta_H$ and $\bar{e}_L = \frac{1}{4-(\frac{\theta_L}{\theta_H})^2}\theta_L$. Therefore, the principal's profit is $E[\bar{\phi}] =$

[8] Revelation principle allows us to focus on truth-telling contracts.

$\frac{1}{4}\theta_H^2 + \frac{1}{2[4-(\frac{\theta_L}{\theta_H})^2]^2}\theta_L^2$, and the agents' payoffs are $E[\bar{\pi}_H] = \frac{2-(\frac{\theta_L}{\theta_H})^2}{2[4-(\frac{\theta_L}{\theta_H})^2]^2}\theta_L^2$ and $E[\bar{\pi}_L] = 0$.

■ **Output monitoring with limited liability**. Here the payment depends only on the output level as well as the agent's reported type, i.e., $p(x, \hat{z})$. Since there are only two types and two output levels, we can denote the corresponding payments to be $p_{1H} \equiv p(1, \theta_H)$, $p_{0H} \equiv p(0, \theta_H)$, $p_{1L} \equiv p(1, \theta_L)$ and $p_{0L} \equiv p(0, \theta_L)$. Therefore, the principal's objective is

$$\max_{p_{1H}, p_{0H}, p_{1L}, p_{0L}, e_H, e_L} \frac{1}{2}[(1-p_{1H})\theta_H e_H - p_{0H}(1-\theta_H e_H) + (1-p_{1L})\theta_L e_L - p_{0L}(1-\theta_L e_L)]$$

subject to the following constraints:

$$e_H = \arg\max_e p_{1H}\theta_H e + p_{0H}(1-\theta_H e) - \frac{e^2}{2}, \qquad e_L = \arg\max_e p_{1L}\theta_L e + p_{0L}(1-\theta_L e) - e^2,$$

$$p_{1H}\theta_H e_H + p_{0H}(1-\theta_H e_H) - \frac{e_H^2}{2} \geq p_{1L}\theta_H e + p_{0L}(1-\theta_H e) - \frac{e^2}{2}, \;\; \forall e$$

$$p_{1L}\theta_L e_L + p_{0L}(1-\theta_L e_L) - e_L^2 \geq p_{1H}\theta_L e + p_{0H}(1-\theta_L e) - e^2, \;\; \forall e$$

$$p_{1H}\theta_H e_H + p_{0H}(1-\theta_H e_H) - \frac{e_H^2}{2} \geq 0, \quad p_{1L}\theta_L e_L + p_{0L}(1-\theta_L e_L) - e_L^2 \geq 0,$$

$$p_{1H},\; p_{0H},\; p_{1L},\; p_{0L} \geq 0.$$

These last conditions are limited liabilities that need to be imposed separately as participation constraints do not automatically guarantee them, unlike in the previous cases. Solving the above problem yields the optimal payments $p_{1H} = p_{1L} = \frac{1}{2}$ and $p_{0H} = p_{0L} = 0$, and the efforts are $e_H = \frac{1}{2}\theta_H$ and $e_L = \frac{1}{4}\theta_L$. Therefore, the principal's profit is $E[\phi] = \frac{1}{8}\theta_H^2 + \frac{1}{16}\theta_L^2$, and the agents' payoffs are $E[\pi_H] = \frac{1}{8}\theta_H^2$ and $E[\pi_L] = \frac{1}{16}\theta_L^2$.[9]

■ **Output monitoring without limited liability**. Solving the output monitoring problem without imposing the limited liability conditions, that is the principal can impose negative payments on the agents, the optimal payments can be derived as $p_{1H} = \frac{2\theta_L^4\theta_H^2 - \theta_L^6 - 8\theta_H^6 + 16\theta_H^4}{16\theta_H^4}$, $p_{0H} = \frac{2\theta_L^4\theta_H^2 - \theta_L^6 - 8\theta_H^6}{16\theta_H^4}$, $p_{1L} = \frac{\theta_L^2(8\theta_H^2 - \theta_L^4)}{16\theta_H^4}$ and $p_{0L} = -\frac{\theta_L^6}{16\theta_H^4}$, and the optimal effort levels are $e_H = \theta_H$ and $e_L = \frac{\theta_L^3}{4\theta_H^2}$. The principal's profit is $E[\phi] = \frac{1}{4}\theta_H^2 + \frac{\theta_L^2}{16\theta_H^2}\theta_L^2$, and the agents' payoffs are $E[\pi_H] = \frac{2\theta_L^4\theta_H^2 - \theta_L^6}{16\theta_H^4}$ and $E[\pi_L] = 0$. In addition, one can find that the principal induces the high type agent to put in the first-best effort, but the payments to both types of agents when output is 0 are negative, that is, the principal needs to ask the agents to make positive transfers if the outcome is bad.

■ **Comparison**. Comparing principal's profit, we find that with limited liability, effort monitoring is better than contribution monitoring, and contribution monitoring is better than output monitoring; this ranking does not change if limited liability is *not* imposed.[10]

[9] Here, both low and high productive agents earn positive rents, as Lawarrée and Audenrode's (1996) Proposition III predicted.

[10] In fact, under effort monitoring, contribution monitoring, and output monitoring without limited liability the principal induces the first-best effort from the high type agent and extracts all the surplus from the low type agent, whereas under output monitoring with limited liability the effort put in by the high type agent is less than the first best and both agents receive positive information rent.

By making production stochastic we want to clarify how our analysis of monitoring goes beyond that of Khalil and Lawarrée. In their model input monitoring refers to the verifiability of agent's effort, thus practically doing away with the moral hazard problem. Alternatively, the principal could rely on output monitoring where output is given by a deterministic production technology. With stochastic output in our formulation, an agent's *contribution* (as opposed to effort) is the combined impact of his type (or ability) and effort without the random error. So our interpretation of contribution, which matches with McAfee and McMillan's concept, is more like Khalil and Lawarrée's *output*. To be able to monitor agent's contribution, thus defined, our principal has to somehow identify the realization of the random error component.[11] Supposing the principal could do this, then the issue analyzed in the above example (as well as our formal model) is whether the principal should monitor by availing error-free information of agent's contribution or the gross output that also has the random component added. What the examples show is that with limited liability, the principal should strictly prefer to monitor the agent's contribution, leaving out the *luck* factor from incentives.[12] Since Khalil and Lawarrée's output is not comparable to our more comprehensive definition of output, we cannot directly compare their results with our formal results later. However, still, the parallel that we can draw between our examples and Khalil and Lawarrée's result is that the principal should avail of the more *refined information* of agent contribution.

A second objective of these single agent examples is that when we turn to team problems, we will see that limited liability does have an impact on how the principal ranks different types of monitoring. So studying the question of monitoring in teams is of independent merit beyond standard principal-agent models.

In the following table we summarize which type of monitoring the principal should favor in the different setups – Khalil and Lawarrée's work, our single agent examples, and the analysis of team problems in McAfee and McMillan and this paper. The table will become easier to follow after one reads the team analysis in the later sections.

Table 1: Choice of monitoring

<table>
<tr><th></th><th></th><th>Effort vs. Contribution</th><th>Contribution vs. Output</th></tr>
<tr><td rowspan="2">Single</td><td>w/o LL</td><td rowspan="2">Effort
(Khalil & Lawarrée; our examples)</td><td>Contribution (our examples)</td></tr>
<tr><td>LL</td><td>Contribution (our examples)</td></tr>
<tr><td rowspan="2">Team</td><td>w/o LL</td><td></td><td>Equivalent (McAfee & McMillan)</td></tr>
<tr><td>LL</td><td></td><td>Contribution (ours)</td></tr>
</table>

[11]This might be possible if the agent's performance is subject to shocks that are commonly observed in the aggregate. For instance, an insurance agent's performance can be assessed in relation to the state of the industry, whether in upswing or downturn.

[12]That luck should not enter into rewards is not a forgone conclusion. Bertrand and Mullainathan (2001) observe that better governed firms pay their CEOs for luck.

3 The Model

A principal wants a team of n agents to undertake a project.[13] Both the principal and the agents are risk neutral. Each agent i is endowed with ability level z_i, $i = 1, ..., n$. There are m different ability levels ranging from the lowest to the highest, i.e., from θ_1 to θ_m, and $m \geq 2$. It is common knowledge that z_i is identically and independently distributed, and for each agent i, $\Pr(z_i = \theta_j) = q_j > 0$ $\forall j$. Only agent i knows his true ability z_i. Each agent i chooses an effort level privately. Let $y_i \in \mathcal{R}_+$ represent agent i's individual contribution which is a combined outcome of i's ability and effort. The agent's cost of contribution is $c(y_i, z_i) \geq 0$, $\forall y_i, z_i$, with $c(y_i, z_i)$ differentiable in y_i. Assume that $c(0, z_i) = 0$ and $c_y \geq 0$ $\forall z_i$, and for any value of $y_i \neq 0$, $c(y_i, \theta_1) > c(y_i, \theta_2) > ... > c(y_i, \theta_m)$. Also, $\forall z_i$, $c(y_i, z_i) \to \infty$ as $y_i \to \infty$. Let $z = (z_1, ..., z_n)$, $z_{-i} = (z_1, ..., z_{i-1}, z_{i+1}, ..., z_n)$, $y = (y_1, ..., y_n)$, and $y_{-i} = (y_1, ..., y_{i-1}, y_{i+1}, ..., y_n)$. The output level x, which is observed by the principal, depends on each one of the team members' contributions and some noise. Assume that the possible output level $x \in X$, where X is a non-empty finite set. Denote the conditional probability of the output x given y by $f(x|y)$, which is uniformly bounded below by some number $a > 0$, i.e., $f(x|y) \geq a > 0$, $\forall x, y$. Also, $f(x|y)$ is continuous in y. The principal's utility is $U(x, z)$ which depends not only on the realized output, but also on the agents' abilities. We also normalize the value of each agent's outside option to be 0.[14]

For any function $\Psi(x, z)$ or $\Psi(y, z)$ or $\Psi(x, y, z)$, let

$$E_{-i}\Psi(\cdot, z) = \sum_{z_1=\theta_1}^{\theta_m} ... \sum_{z_{i-1}=\theta_1}^{\theta_m} \sum_{z_{i+1}=\theta_1}^{\theta_m} ... \sum_{z_n=\theta_1}^{\theta_m} \Psi(\cdot, z)\Big[\prod_{\ell \neq i} \Pr(z_\ell)\Big];$$

$$E_z\Psi(\cdot, z) = \sum_{z_1=\theta_1}^{\theta_m} ... \sum_{z_n=\theta_1}^{\theta_m} \Psi(\cdot, z)\Big[\prod_{\ell=1}^{n} \Pr(z_\ell)\Big];$$

$$E_x\Psi(x, \cdot) = \sum_{x} \Psi(x, z) \cdot f(x|y).$$

By the Revelation Principle, we will focus on truth-telling mechanisms. Let $\hat{z}_i$ denote agent i's report to the principal about his ability, and $\hat{z}$ and $\hat{z}_{-i}$ are the reported type profiles with similar meaning as before.

Contract specification, type reporting and subsequent contribution choices proceed in the following order.

Stage 0: Nature chooses the agents' abilities (or types). Each agent learns his type.

Stage 1: The principal chooses the monitoring mechanism (contributions or output) and offers each agent a contract specifying the wage contingent on the profile of declared types and the principal's verification of the agent's contribution or the team's output.

Stage 2: The agents publicly declare their types simultaneously, and each agent signs the

[13] Although we work in a team setting throughout the paper, our results are applicable in a single agent setting.

[14] The single agent formulation in section 2 does not fall into our setup exactly, as it violates some of the assumptions here. For instance, there, $f(x|0) = 0$, the value of y_i is restricted to be less than 1. However, these minor differences do not affect the essential parts of our arguments. We impose relatively strong assumptions here only to prove the main result in a general environment.

contract corresponding to his announced type so long as it guarantees at least his reservation utility.

Stage 3: The agents choose their contribution levels simultaneously, determining the output.

Terminal Stage: Either individual contributions or the output level is revealed to the principal depending on the monitoring mechanism in place. Wages are paid according to the signed contract. ||

The real actions happen in stages 1-3, with no actions but routine contract background preparation and contract fulfillment occurring in stage 0 and the terminal stage. The resulting extensive game we call Γ.

The equilibrium concept of the extensive form game Γ is Perfect Bayesian Equilibrium (or *PBE*). Since we are considering a direct revelation game and there is no way for the agents to verify whether any of the other agents have chosen off-equilibrium actions (i.e., non-truthfully reported types), we will assign any agent i's belief as follows:

$$\mu_i(z_{-i}) = \begin{cases} 1 & \text{if} \quad z_{-i} = \hat{z}_{-i} \\ 0 & \text{if} \quad z_{-i} \neq \hat{z}_{-i}. \end{cases} \tag{1}$$

During the game the agents will proceed to calculate their payoffs both on- and off-the-equilibrium path (following an agent's own deviation while reporting his type) using the above beliefs. By *equilibrium* we will always mean *PBE.* Thus, the game will be solved in the standard way using backwards induction.

■ **Output Monitoring Contract.** In the output monitoring contract, the principal can only observe the aggregate team output, so he commits in advance to a payment rule $p_i(x, \hat{z}_i, \hat{z}_{-i})$. That is, each agent's payment depends on the realized output, his report of his own ability (or type), and the other agents' declared types.

The principal's profit and the agents' payoffs are respectively

$$\phi(x, z, \hat{z}) = U(x, z) - \sum_{i=1}^{n} p_i(x, \hat{z}_i, \hat{z}_{-i}),$$
$$\pi_i(x, z_i, \hat{z}_i, y_i, \hat{z}_{-i}) = p_i(x, \hat{z}_i, \hat{z}_{-i}) - c(y_i, z_i), \quad i = 1, ..., n.$$

Thus, the principal solves the following program:

$$[\mathcal{P}_{\text{out}}] \qquad \max_{\{p_i(\cdot), y_i(\cdot)\}} E_z \sum_x \left[U(x, z) - \sum_{i=1}^{n} p_i(x, z_i, z_{-i})\right] f(x|y(z))$$

subject to the feasibility constraints:

$$\sum_x p_i(x, z_i, z_{-i}) f(x|y(z_i, z_{-i})) - c(y_i(z_i, z_{-i}), z_i)$$
$$\geq \sum_x p_i(x, z_i, z_{-i}) f(x|y'_i, y_{-i}(z_i, z_{-i})) - c(y'_i, z_i), \quad \forall i, z_i, z_{-i}, y'_i \tag{IC-con(out)}$$

$$E_{-i}\Big[\sum_x p_i(x,z_i,z_{-i})f(x|y(z_i,z_{-i}))-c(y_i(z_i,z_{-i}),z_i)\Big]$$
$$\geq E_{-i}\Big[\sum_x p_i(x,\hat{z}_i,z_{-i})f(x|y_i'(z_i,\hat{z}_i,z_{-i}),y_{-i}(\hat{z}_i,z_{-i}))-c(y_i'(z_i,\hat{z}_i,z_{-i}),z_i)\Big],\quad \forall i,z_i,\hat{z}_i,y_i'(z_i,\hat{z}_i,z_{-i}) \tag{IC-type(out)}$$
$$E_{-i}\Big[\sum_x p_i(x,z_i,z_{-i})f(x|y(z_i,z_{-i}))-c(y_i(z_i,z_{-i}),z_i)\Big]\geq 0,\quad \forall i,z_i \tag{PC(out)}$$
$$p_i(x,\hat{z}_i,\hat{z}_{-i})\geq 0,\quad \forall i,x,\hat{z}_i,\hat{z}_{-i}. \tag{LL(out)}$$

where $y_i'(z_i,\hat{z}_i,z_{-i})$ is any arbitrary choice of contribution by agent i with type z_i.

Different from McAfee and McMillan's (1991) analysis we impose the limited liability constraints for the agents as in LL(out), i.e., regardless of the output level or type declarations the agents cannot be asked to make positive transfers to the principal. In reality, the agents might be financially constrained to make upfront payment feasible.

Denote $p(.)=(p_1(.),...,p_n(.))$. The following result gives the optimal monitoring question a proper benchmark.

Lemma 1 *An optimal output monitoring contract solving the program $[\mathcal{P}_{out}]$ always exists.*

Denote $C^*_{out}\equiv\Big\{(p^*(\cdot),y^*(\cdot)) : (p^*(\cdot),y^*(\cdot))$ solves the program $[\mathcal{P}_{out}]\Big\}$ to be the set of optimal output monitoring contracts. Thus, for any $(p^*(x,\hat{z}_i,\hat{z}_{-i}),y^*(\hat{z}_i,\hat{z}_{-i}))\in C^*_{out}$, when other players truthfully report their types z_{-i} and choose contributions according to $y^*_{-i}(\hat{z}_i,z_{-i})$, the interim expected payoff of a type z_i agent reporting $\hat{z}_i$ and contributing y_i is

$$E_x\pi_i(x,z_i,\hat{z}_i,y_i,z_{-i})=\sum_x p_i^*(x,\hat{z}_i,z_{-i})f(x|y_i,y^*_{-i}(\hat{z}_i,z_{-i}))-c(y_i,z_i),\quad \forall z_{-i}.$$

Define
$$\hat{Y}_i^{out}(z_i,\hat{z}_i,z_{-i})\equiv\big\{\arg\max_{y_i} E_x\pi_i(x,z_i,\hat{z}_i,y_i,z_{-i})\big\},$$

where $\hat{z}_i\neq z_i$, i.e., given other agents' type profile z_{-i}, any $\hat{y}_i^{out}(z_i,\hat{z}_i,z_{-i})\in\hat{Y}_i^{out}(z_i,\hat{z}_i,z_{-i})$ is the optimal contribution level of agent i if he misreports his type to be $\hat{z}_i$. That $\hat{Y}_i^{out}(z_i,\hat{z}_i,z_{-i})\neq\emptyset$ can be guaranteed by using an argument similar to the last part of Step 2 in the proof of Lemma 1.

Clearly, by design, the highest possible ex-ante expected payoff to agent i if he misreports his type as $\hat{z}_i$ cannot be higher than what he could obtain by truthful reporting, that is:

$$E_{-i}\Big[\sum_x p_i^*(x,z_i,z_{-i})f(x|y^*(z_i,z_{-i}))-c(y_i^*(z_i,z_{-i}),z_i)\Big]$$
$$\geq E_{-i}\Big[\sum_x p_i^*(x,\hat{z}_i,z_{-i})f(x|\hat{y}_i^{out}(z_i,\hat{z}_i,z_{-i}),y^*_{-i}(\hat{z}_i,z_{-i}))-c(\hat{y}_i^{out}(z_i,\hat{z}_i,z_{-i}),z_i)\Big],\quad \forall i,z_i,\hat{z}_i. \tag{IC-type*(out)}$$

This inequality will be useful later on.

■ **Contributions Monitoring Contract.** Now, suppose the principal could costlessly monitor each individual's contribution, so that he could pay agent i according to his contribution y_i as well as the reported profile of abilities $(\hat{z}_i, \hat{z}_{-i})$. Denote the payment by $\bar{p}_i(y_i, \hat{z}_i, \hat{z}_{-i})$.[15]

The principal's profit and the agents' payoffs are respectively

$$\bar{\phi}(x, y, z, \hat{z}) = U(x, z) - \sum_{i=1}^{n} \bar{p}_i(y_i, \hat{z}_i, \hat{z}_{-i}),$$

$$\bar{\pi}_i(z_i, \hat{z}_i, y_i, \hat{z}_{-i}) = \bar{p}_i(y_i, \hat{z}_i, \hat{z}_{-i}) - c(y_i, z_i), \quad i = 1, ..., n.$$

For a given payment function $\bar{p}_i(y_i, \hat{z}_i, \hat{z}_{-i})$, when other players truthfully report their types, the interim payoff of agent i with type z_i reporting $\hat{z}_i$ and contributing y_i is

$$\bar{\pi}_i(z_i, \hat{z}_i, y_i, z_{-i}) = \bar{p}_i(y_i, \hat{z}_i, z_{-i}) - c(y_i, z_i), \quad \forall z_{-i}.$$

Thus, the principal solves the following program:

$$[\mathcal{P}\text{con}] \qquad \max_{\{\bar{p}_i(\cdot), \bar{y}_i(\cdot)\}} E_z \sum_x \left[U(x, z) - \sum_{i=1}^{n} \bar{p}_i(\bar{y}_i(z_i, z_{-i}), z_i, z_{-i})\right] f(x|\bar{y}(z_i, z_{-i}))$$

subject to the feasibility constraints:

$$\bar{\pi}_i(z_i, z_i, \bar{y}_i(z_i, z_{-i}), z_{-i}) \geq \bar{\pi}_i(z_i, z_i, y_i, z_{-i}), \quad \forall i, z_i, z_{-i}, y_i \qquad \text{(IC-con(con))}$$

$$E_{-i}\bar{\pi}_i(z_i, z_i, \bar{y}_i(z_i, z_{-i}), z_{-i}) \geq E_{-i}\bar{\pi}_i(z_i, \hat{z}_i, y_i(z_i, \hat{z}_i, z_{-i}), z_{-i}), \quad \forall i, z_i, \hat{z}_i, y_i(z_i, \hat{z}_i, z_{-i}) \qquad \text{(IC-type(con))}$$

$$E_{-i}\bar{\pi}_i(z_i, z_i, \bar{y}_i(z_i, z_{-i}), z_{-i}) \geq 0, \quad \forall i, z_i \qquad \text{(PC(con))}$$

$$\bar{p}_i(y_i, \hat{z}_i, \hat{z}_{-i}) \geq 0, \quad \forall i, y_i, \hat{z}_i, \hat{z}_{-i}, \qquad \text{(LL(con))}$$

where $y_i(z_i, \hat{z}_i, z_{-i})$ is any arbitrary choice of contribution by agent i.

4 Contributions vs. Output Monitoring

■ **Payoff Equivalent Contributions Monitoring**

Definition 1 (Punishing contract) *A punishing contract $\mathcal{M}_{con}$ consists of a target level of contribution $y_i^c(\hat{z}_i, \hat{z}_{-i})$ and a payment function $\check{p}_i(y_i, \hat{z}_i, \hat{z}_{-i})$ such that:*

$$\check{p}_i(y_i, \hat{z}_i, \hat{z}_{-i}) = \begin{cases} r(\hat{z}_i, \hat{z}_{-i}), & \text{if } y_i = y_i^c(\hat{z}_i, \hat{z}_{-i}) \\ 0, & \text{otherwise.} \end{cases} \qquad (2)$$

[15] Note that the payment to agent i is restricted to depend only on i's contribution. Admittedly this weakens the principal's hand but given that ultimately we are going to show dominance of contributions monitoring, allowing a more general payment function that depends on other agents' contributions as well would retain the dominance result if not strengthen it further.

That is, the principal specifies a contribution level so that for contributions differing from the specified level agent i is penalized heavily.[16] So the agent should either choose to meet the target $y_i^c(\hat{z}_i, \hat{z}_{-i})$, or completely shirk by contributing 0.

Proposition 1 (Simple contract) *The equilibrium outcome of any feasible contributions monitoring contract can be replicated by a simple punishing contract as in Definition 1.*

From here onwards we need to consider only the punishing contract in the class of contributions monitoring contract. Define

$$\hat{Y}_i^{con}(z_i, \hat{z}_i, z_{-i}) \equiv \left\{ \arg\max_{y_i} \bar{\pi}_i(z_i, \hat{z}_i, y_i, z_{-i}) \right\},$$

where $\hat{z}_i \neq z_i$, i.e., given other agents' type profile z_{-i}, any $\hat{y}_i^{con}(z_i, \hat{z}_i, z_{-i}) \in \hat{Y}_i^{con}(z_i, \hat{z}_i, z_{-i})$ is the optimal contribution level of agent i if he misreports his type to be $\hat{z}_i$. Also, we know that $\hat{Y}_i^{con}(z_i, \hat{z}_i, z_{-i}) \neq \emptyset$ as noted above (before Proposition 1). Thus, the IC-type(con) constraint can be replaced by the following:

$$E_{-i}\bar{\pi}_i(z_i, z_i, \bar{y}_i(z_i, z_{-i}), z_{-i}) \geq E_{-i}\bar{\pi}_i(z_i, \hat{z}_i, \hat{y}_i^{con}(z_i, \hat{z}_i, z_{-i}), z_{-i}), \quad \forall i, z_i, \hat{z}_i. \qquad \text{(IC-type*(con))}$$

Proposition 2 (Payoff equivalence) *Given any feasible output monitoring mechanism* $\{p_i'(x, \hat{z}_i, \hat{z}_{-i}), y_i'(\hat{z}_i, \hat{z}_{-i})\}$ *specified in the program* $[\mathcal{P}_{out}]$*, the contributions monitoring contract with target contributions* $y_i^c(\hat{z}_i, \hat{z}_{-i}) = y_i'(\hat{z}_i, \hat{z}_{-i})$ *and payment function*

$$\breve{p}_i(y_i, \hat{z}_i, \hat{z}_{-i}) = \begin{cases} \sum_x p_i'(x, \hat{z}_i, \hat{z}_{-i}) f(x | y'(\hat{z}_i, \hat{z}_{-i})), & \text{if } y_i = y_i'(\hat{z}_i, \hat{z}_{-i}) \\ 0, & \text{otherwise} \end{cases} \qquad (3)$$

for each i, is feasible, induces the same contributions and generates the same interim and ex-ante expected profit and payoffs for the principal and the agents (for each of their types) in equilibrium, as in the given output monitoring contract.

Payoff equivalence works as follows. Given any feasible output based contract, the principal could calculate the equilibrium individual contribution and the interim expected payoff for each type of each agent. Thus, in the situation where each agent's individual contribution can be perfectly monitored, the principal could induce the same profile of contributions so as to maintain the same expected output, by promising each agent the same interim payoff (as that under output monitoring) if the agent meets the target. Thus, the principal as well as the agents' ex-ante expected equilibrium payoff will be the same as before. On the other hand, if the agent deviates to report as another type under output monitoring, he has the freedom to choose his deviation contribution, whereas under contributions monitoring the deviation contribution level is specified by the principal. If the agent does not meet the target, his contribution yields no reward. Such deviation contribution level chosen by the principal may not necessarily coincide with the agent's interest. Therefore, we can see that indeed agents have (weakly) less incentives to deviate under contributions monitoring.

[16] A more natural mechanism would be to give the non-negative reward so long as i's contribution is at least $\bar{y}_i(\hat{z}_i, \hat{z}_{-i})$ and zero reward otherwise. All our analysis will hold for this alternative mechanism.

■ **Strict Dominance of Contributions Monitoring.** To show strict dominance of contributions monitoring, we first present two useful properties, one each for output and contributions monitoring contracts.

Proposition 3 *For any feasible output monitoring contract* $\{p_i(x, \hat{z}_i, \hat{z}_{-i}), y_i(\hat{z}_i, \hat{z}_{-i})\}$ *specified in the program* $[\mathcal{P}_{out}]$*, for any agent* i*, either every type of him earns* 0 *information rent, or every type of him earns strictly positive information rent.*

Thus, we know that for the optimal output monitoring contract, it must be true that either all types of all agents get zero information rent or for at least one agent, every type of him earns strictly positive information rent. Due to limited liability, the former implies that $p_i(x, z_i, z_{-i}) = 0$ and $y_i(z_i, z_{-i}) = 0\ \forall i, x, z_i, z_{-i}$, which we call a *null contract*. This is because if $p_i(x, z_i, z_{-i}) > 0$ for some (x, z_i, z_{-i}), the agent can always put in zero contribution in all contingencies and earn a positive rent, whereas if $y_i(z_i, z_{-i}) > 0$ for some z (and $p_i(x, z_i, z_{-i}) = 0$ for all (x, z)) then agent i's payoff will be negative. In fact, theoretically the null contract might even be optimal under certain conditions, e.g., when $c(y_i, z_i)$ is very large $\forall y_i \neq 0$. But it is inconceivable that a principal will hire a group of agents for nothing. To be more relevant for organizational design, for the optimal output monitoring contract we assume that for at least one agent, every type of him earns strictly positive information rent. In the rest of the analysis, we will consider only this case. We therefore re-state Proposition 3 as follows:

Proposition 3′ (Information rent in optimal output monitoring) *For any optimal output monitoring contract* $\{p_i^*(x, \hat{z}_i, \hat{z}_{-i}), y_i^*(\hat{z}_i, \hat{z}_{-i})\}$ *solving the program* $[\mathcal{P}_{out}]$*, there is at least one agent such that every type of him earns strictly positive information rent.*

The following lemma will be useful to introduce the property of the optimal contributions monitoring contract.

Lemma 2 *For any punishing contract* $\mathcal{M}_{con}$ *that is feasible, suppose there exists a particular type of an agent, not necessarily his lowest type, such that none of his IC-type*(con) and PC(con) constraints are binding. Then, there exists another feasible punishing contract which generates a strictly higher expected profit for the principal.*

The proof of Lemma 2 in the Appendix uses the following logic. With non-binding incentive and participation constraints, the particular type (of the agent) must be receiving a positive reward for some contingency. Now the principal can lower (only) this reward (leaving all other payments unchanged) appropriately without violating any of the IC-type*(con), PC(con) and LL(con) constraints. This new contract will be feasible and none of the contributions (of this or any other agent) will change. Thus, doing so leads to a higher expected profit for the principal.

By making use of Lemma 2, we will derive a property for the optimal punishing contract. But before that we need the following assumption:

Assumption 1 *An optimal contributions monitoring contract solving the program* $[\mathcal{P}_{con}]$ *always exists.*

Unlike Lemma 1 where we established the existence of optimal output monitoring contract, we need Assumption 1. Showing existence of the optimal contributions monitoring contract is a difficult exercise given that the principal will have to solve an optimization program in the space of functions $\{\bar{p}_i(\bar{y}_i, z_i, z_{-i}), \bar{y}_i(z_i, z_{-i})\}$ where $\bar{y}_i \in [0, \infty)$.[17]

Given Assumption 1, an optimal punishing contract solving the program $[\mathcal{P}_{\text{con}}]$ will exist.

Proposition 4 (Information rent in contributions monitoring) *Suppose Assumption 1 holds. Then in the optimal punishing contract, for each agent* i*, his lowest type must earn zero information rent, i.e.,* $E_{-i}\bar{\pi}_i(\theta_1, \theta_1, \bar{y}_i(\theta_1, z_{-i}), z_{-i}) = 0$*,* $\forall i$*.*

By Proposition 2, we could replicate the optimal output monitoring contract with a punishing contract. By Proposition 3′, we know that under the optimal output monitoring contract and the replicated punishing contract, for at least one agent, the lowest type of him earns strictly positive information rent. However, by Proposition 4, we know that such punishing contract cannot be optimal since the lowest type must earn *zero* ex-ante expected payoff under the optimal punishing contract. Thus, we have our main result as follows.

Proposition 5 (Strict dominance) *The contributions monitoring contract is strictly superior to output monitoring contract, i.e., the principal's ex-ante expected profit from monitoring individual contributions is strictly higher than that when he observes only the total output.*

The above result is quite different from that of McAfee and McMillan. Theirs is about the equivalence of two monitoring mechanisms, whereas we show strict dominance of contributions monitoring. Technically, LL(out) requires $p_i(x, \hat{z}_i, \hat{z}_{-i}) \geq 0$ while LL(con) requires $\breve{p}_i(y_i, \hat{z}_i, \hat{z}_{-i}) = \sum_x p_i(x, \hat{z}_i, \hat{z}_{-i}) f(x|y^c(\hat{z}_i, \hat{z}_{-i})) \geq 0$ when $y_i = y_i^c(\hat{z}_i, \hat{z}_{-i})$. As $\breve{p}_i(y_i^c(\hat{z}_i, \hat{z}_{-i}), \hat{z}_i, \hat{z}_{-i})$ is constructed in expected form, LL(con) is a less stringent condition than LL(out). Thus, under limited liability punishing contracts embrace more possibilities (where some $p_i(x, \hat{z}_i, \hat{z}_{-i})$'s can even be negative-valued) than is possible under output monitoring. This makes the punishing contract more powerful than the output monitoring contract, yielding a beneficial improvement for the principal.

■ **Where exactly does McAfee-McMillan's contributions monitoring mechanism miss out?** In McAfee-McMillan's output monitoring mechanism, due to the absence of the limited liability restriction the lowest type of every agent could be pushed to receive zero information rent.[18] Thus when one replicates the outcome under optimal output monitoring contract by a punishing contract, the lowest type agents receive zero rent in the replicated contract, which coincides with the property of the optimal punishing contract without limited

[17] It might still be possible to prove existence by showing that an optimal punishing contract exists following the same method as in Lemma 1, given that the set of target contributions, $\{\breve{y}_i^c(z_i, z_{-i})\}$, is finite.

[18] In our model, as shown in Proposition 3′, after including limited liability, it is not possible to extract all the surplus from the lowest type of every agent. To maintain incentives for high contributions, it usually requires a big pay dispersion based on different levels of output, and to minimize the cost to the principal negative transfer for low output is thus necessary. However, with the restriction of limited liability, the minimum payment can only be set to 0, which makes the creation of the spread of payments more costly for the principal.

liability.[19] Then the principal might not have any more room to improve beyond the replicating punishing contract. With limited liability brought in, the optimal output monitoring contract ties the principal's hands more severely than the punishing contract, enabling a strict improvement by the latter mechanism.

■ **Ranking of Monitoring Mechanisms: Single Agent vs. Teams.** As we had earlier indicated in section 2, ranking of monitoring in team problems is somewhat different from that in single agent contracting. For teams limited liability makes a subtle difference as it changes from strict dominance of monitoring of individual contributions over output monitoring (under LL) to equivalence (in the absence of LL), in contrast to uniformly strict dominance of contributions monitoring in the single agent examples provided earlier. What makes output monitoring to catch up with contributions monitoring in the case of teams as one drops the requirement of limited liability? Under contributions monitoring the principal can identify who are the failing agents, so he does not have to penalize all team members uniformly while such punishment would be almost necessary for low performing teams under output monitoring. Thus, as the limited liability requirement is relaxed, output monitoring is likely to empower the principal much more than contributions monitoring, hence the equivalence result of McAfee and McMillan. In the single agent case on the other hand, the principal experiences only a "small" improvement in incentives flexibility under output monitoring (as one moves from limited liability to no limited liability), thus leaving strict superiority of contribution monitoring unchanged. This (i.e., small improvement) is because the principal now has to decide whether low performance is due to the single agent's non-diligence or pure bad luck, as opposed to having to identify who among the team members are delinquent. We can thus see how team monitoring poses a different hurdle than in the principal-agent models. The ranking of monitoring mechanisms can be summarized as follows:

Table 2: Ranking of monitoring

	Single agent (example)	Team
LL	Contribution $\succ$ Output	Contribution $\succ$ Output
w/o LL	Contribution $\succ$ Output	Contribution $\sim$ Output

A Appendix

Proof of Lemma 1. First, we can see that the contract with the payment function $p_i(x, \hat{z}) = 0$, $\forall i, x, \hat{z}$ and equilibrium contribution level $y_i(\hat{z}) = 0$, $\forall i, \hat{z}$ satisfies all the feasibility constraints. Let $\tilde{V}$ denote the principal's expected profit under such contract.

Next, we are going to show the existence of optimal solution.

Since the domain of $p_i(x, \hat{z})$ and $y_i(\hat{z})$ are finite, choosing the set of such functions will be the same as choosing a finite number of vectors. Let V be the space of $n \times m^n \times (1 + |X|)$

[19] Our Proposition 4 (derived under the limited liability restriction) also applies to the case without limited liability.

dimensional vectors which satisfy all the feasibility constraints and yields principal the expected payoff at least $\tilde{V}$. Thus, each element $v_k \in V$ can be represented as

$$\begin{aligned} v_k(x,\hat{z}) &= (y_k(\hat{z}), p_k(x,\hat{z})) \\ &= (y_{1k}(\hat{z}), y_{2k}(\hat{z}), ..., y_{nk}(\hat{z}), p_{1k}(x,\hat{z}), p_{2k}(x,\hat{z}), ..., p_{nk}(x,\hat{z})). \end{aligned}$$

Claim 1: The objective function in the program $[\mathcal{P}_{out}]$ *is continuous on* V.

This is trivial since the objective function only consists of simple arithmetic operators, so the proof is omitted here.

Claim 2: V *is bounded.*

Since there are finite number of output x and type profile z, there always exists $\bar{x}$ and $\bar{z}$ such that $U(x,z) \leq U(\bar{x},\bar{z})$, $\forall x,z$. As $p_i(x,z) \geq 0$, $\forall i,x,z$, the principal's maximum possible expected profit is $U(\bar{x},\bar{z})$. Since $\Pr(z_i = \theta_j) = q_j > 0$ $\forall j$, there always exists a $\underline{q} \in \{q_1, q_2, ..., q_m\}$ such that $q_j \geq \underline{q}$. Thus, the principal will choose a payment function such that $p_i(x,z) \leq \frac{U(\bar{x},\bar{z})-\tilde{V}}{a\underline{q}^n}$ $\forall i,x,z$. Otherwise, his expected profit will be strictly smaller than $\tilde{V}$. By LL(out), we know that $p_i(x,z) \geq 0$ $\forall i,x,z$. Thus, the choice set of the function $p_i(x,z)$ is bounded.

Since $c(y_i,z_i) \to \infty$ as $y_i \to \infty$, $p_i(x,z)$ is bounded uniformly by $\frac{U(\bar{x},\bar{z})-\tilde{V}}{a\underline{q}^n}$ and $0 < f(x|y) \leq 1$, we have $\sum_x p_i(x,z)f(x|y) - c(y_i,z_i) \to -\infty$ as $y_i \to \infty$. Thus, for every type profile z, there exists a value $\bar{y}_i(z)$ such that the unconstrained maximization problem $\max_{y_i} \sum_x p_i(x,z)f(x|y) - c(y_i,z_i)$ is the same as if it is solved with the addition of the constraint $y_i(z) \leq \bar{y}_i(z)$. Also, $y_i(z) \geq 0$ $\forall i,z$. Thus, given any particular value of z, the choice set of y_i of this maximization problem is closed and bounded. Also, it is easily seen that $\sum_x p_i(x,z)f(x|y) - c(y_i,z_i)$ is continuous in y_i. Thus, the maximizer(s) of this problem exists. Thus, the choice set of the function $y_i(z)$ of the original problem is bounded.

Claim 3: V *is closed.*

The proof is also trivial since all the constraints are weak inequalities.

Claim 4: The optimal solution exists.

Since V is closed and bounded, V is compact. Since the objective function is continuous on V, the optimal output monitoring contract exists. **Q.E.D.**

Proof of Proposition 1. Fix any feasible contributions monitoring contract $\{\bar{p}_i(y_i,\hat{z}_i,\hat{z}_{-i}), \bar{y}_i(\hat{z}_i,\hat{z}_{-i})\}$. Consider the proposed punishing contract with the target level of contribution $y_i^c(\hat{z}_i,\hat{z}_{-i}) = \bar{y}_i(\hat{z}_i,\hat{z}_{-i})$ and payment function

$$\breve{p}_i(y_i,\hat{z}_i,\hat{z}_{-i}) = \begin{cases} \bar{p}_i(\bar{y}_i(\hat{z}_i,\hat{z}_{-i}),\hat{z}_i,\hat{z}_{-i}), & \text{if} \quad y_i = \bar{y}_i(\hat{z}_i,\hat{z}_{-i}) \\ 0, & \text{otherwise.} \end{cases}$$

Clearly, $\breve{p}_i(\bar{y}_i(\hat{z}_i,\hat{z}_{-i}),\hat{z}_i,\hat{z}_{-i}) \geq 0$ $\forall i,\hat{z}_i,\hat{z}_{-i}$ since $\bar{p}_i(y_i,\hat{z}_i,\hat{z}_{-i}) \geq 0$ $\forall i,y_i,\hat{z}_i,\hat{z}_{-i}$. Thus, LL(con) is satisfied.

Also, IC-con(con) is satisfied, because if agent i truthfully reports his type z_i, and contributes $\bar{y}_i(z_i,z_{-i})$, his interim payoff is $\bar{p}_i(\bar{y}_i(z_i,z_{-i}),z_i,z_{-i}) - c(\bar{y}_i(z_i,z_{-i}),z_i) = \bar{\pi}_i(z_i,z_i,\bar{y}_i(z_i,z_{-i}),z_{-i})$ $\geq \bar{\pi}_i(z_i,z_i,y_i,z_{-i}) = \bar{p}_i(y_i,z_i,z_{-i}) - c(y_i,z_i) \geq 0 - c(y_i,z_i)$ $\forall y_i$.

Furthermore, since his interim payoff under the punishing contract, following truthful reporting (i.e., $\hat{z}_i = z_i$), is the same as that under the given feasible contributions monitoring contract, his ex-ante expected payoff under truthful reporting will also be the same. Thus, PC(con) is satisfied.

Lastly, when agent i deviates to report as type $\hat{z}_i \neq z_i$, for any type profile z_{-i}, if he contributes $\bar{y}_i(\hat{z}_i, z_{-i})$, his interim payoff under the given contributions monitoring contract is the same as that under the punishing contract: $\bar{p}_i(\bar{y}_i(\hat{z}_i, z_{-i}), \hat{z}_i, z_{-i}) - c(\bar{y}_i(\hat{z}_i, z_{-i}), z_i)$. If he chooses any other contribution level $y_i' \neq \bar{y}_i(\hat{z}_i, z_{-i})$, his interim payoff under the given contributions monitoring contract is

$$\bar{p}_i(y_i', \hat{z}_i, z_{-i}) - c(y_i', z_i),$$

whereas his payoff under the punishing contract is

$$0 - c(y_i', z_i).$$

Since $\bar{p}_i(y_i', \hat{z}_i, z_{-i}) \geq 0$, the agent's interim payoff under the contributions monitoring contract is weakly higher for all y_i. Thus, his ex-ante expected payoff from any non-truthful reporting is weakly higher under the contributions monitoring contract. Since truthful reporting was incentive compatible under the contributions monitoring contract, IC-type(con) is also satisfied for the punishing contract.

Thus the proposed punishing contract induces the same outcome in the ex ante as well as interim stages as in the contributions monitoring contract. **Q.E.D.**

Proof of Proposition 2. Fix any feasible output monitoring contract $\{p_i'(x, \hat{z}_i, \hat{z}_{-i}), y_i'(\hat{z}_i, \hat{z}_{-i})\}$. Consider the proposed punishing contract with payment function $\breve{p}_i(y_i, \hat{z}_i, \hat{z}_{-i})$ and the target contribution $y_i^c(\hat{z}_i, \hat{z}_{-i}) = y_i'(\hat{z}_i, \hat{z}_{-i})\ \forall i$. Since $p_i'(x, \hat{z}_i, \hat{z}_{-i}) \geq 0, \forall i, x, \hat{z}_i, \hat{z}_{-i}$, we have

$$\sum_x p_i'(x, \hat{z}_i, \hat{z}_{-i}) f(x|y'(\hat{z}_i, \hat{z}_{-i})) \geq 0, \quad \forall i, \hat{z}_i, \hat{z}_{-i}.$$

Thus,

$$\breve{p}_i(y_i, \hat{z}_i, \hat{z}_{-i}) \geq 0, \quad \forall i, y_i, \hat{z}_i, \hat{z}_{-i},$$

i.e., LL(con) is satisfied.

Next, we show that under the proposed punishing contract, each type of each agent's equilibrium individual contribution, his interim and ex-ante expected payoff and the principal's interim and ex-ante expected profit will be the same as those under the given feasible output monitoring mechanism.

Suppose other agents report truthfully. Given the reporting profile $(\hat{z}_i, z_{-i})$, if agent i contributes $y_i'(\hat{z}_i, z_{-i})$, the principal rewards him

$$\sum_x p_i'(x, \hat{z}_i, z_{-i}) f(x|y'(\hat{z}_i, z_{-i})).$$

Otherwise, the agent will get nothing. Suppose agent i truthfully reports his type z_i. If he chooses to contribute $y'_i(z_i, z_{-i})$, his interim payoff is

$$\bar{\pi}_i(z_i, z_i, y'_i(z_i, z_{-i}), z_{-i}) = \sum_x p'_i(x, z_i, z_{-i}) f(x|y'(z_i, z_{-i})) - c(y'_i(z_i, z_{-i}), z_i).$$

If he does not contribute $y'_i(z_i, z_{-i})$, his interim payoff is at most 0. Since $\sum_x p'_i(x, z_i, z_{-i}) f(x|y'(z_i, z_{-i})) - c(y'_i(z_i, z_{-i}), z_i) \geq 0$ (this is satisfied since under output monitoring the agent has the default option of choosing 0 contribution and limited liability guarantees non-negative payment), choosing $y'_i(z_i, z_{-i})$ is optimal.

Since the principal is risk neutral, and he induces each type of each agent i to contribute $y'_i(z_i, z_{-i})$ using the same amount of (expected) payment as that under the output monitoring mechanism $\{p'_i(x, \hat{z}_i, \hat{z}_{-i}), y'_i(\hat{z}_i, \hat{z}_{-i})\}$, his interim and ex-ante expected profits are the same as before. Also, each type of each agent's interim and ex-ante expected payoffs are the same as that under the given output monitoring contract, thus satisfying PC(con).

Now, we are left to show that agent i with true type z_i would truthfully report his type. If the agent is truthful, his ex-ante expected equilibrium payoff under the punishing contract is

$$E_{-i}\bar{\pi}_i(z_i, z_i, y'_i(z_i, z_{-i}), z_{-i}) = E_{-i}[\sum_x p'_i(x, z_i, z_{-i}) f(x|y'(z_i, z_{-i})) - c(y'_i(z_i, z_{-i}), z_i)].$$

If he reports his type as $\hat{z}_i \neq z_i$, his maximum ex-ante expected deviation payoff is

$$\begin{aligned} &E_{-i}[\max\{\breve{p}_i(y'_i(\hat{z}_i, z_{-i}), \hat{z}_i, z_{-i}) - c(y'_i(\hat{z}_i, z_{-i}), z_i), 0\}] \\ =&E_{-i}[\max\{\sum_x p'_i(x, \hat{z}_i, z_{-i}) f(x|y'(\hat{z}_i, z_{-i})) - c(y'_i(\hat{z}_i, z_{-i}), z_i), 0\}]. \end{aligned}$$

Given the IC-type(out) and LL(out), we know

$$\begin{aligned} &E_{-i}[\sum_x p'_i(x, z_i, z_{-i}) f(x|y'(z_i, z_{-i})) - c(y'_i(z_i, z_{-i}), z_i)] \\ \geq &E_{-i}[\max\{\sum_x p'_i(x, \hat{z}_i, z_{-i}) f(x|y'(\hat{z}_i, z_{-i})) - c(y'_i(\hat{z}_i, z_{-i}), z_i), \sum_x p'_i(x, \hat{z}_i, z_{-i}) f(x|0, y'_{-i}(\hat{z}_i, z_{-i}))\}], \quad \forall \hat{z}_i \\ \geq &E_{-i}[\max\{\sum_x p'_i(x, \hat{z}_i, z_{-i}) f(x|y'(\hat{z}_i, z_{-i})) - c(y'_i(\hat{z}_i, z_{-i}), z_i), 0\}], \quad \forall \hat{z}_i\ . \end{aligned}$$

Thus, the truth-telling constraints IC-type(con) are also satisfied. **Q.E.D.**

Proof of Proposition 3. First, we know that a higher type should earn no less than a lower type, since the higher type can always mimic as the lower type and he has (weakly) lower cost. Thus, if the lowest type of agent i earns strictly positive information rent, the higher type of him must also earn strictly positive information rent.

Next, suppose for a feasible output monitoring contract $\{p_i(x, z_i, z_{-i}), y_i(z_i, z_{-i})\}$, there exists an agent i such that the lowest type of him earns 0 information rent, i.e.,

$$E_{-i}[\sum_x p_i(x, \theta_1, z_{-i}) f(x|y(\theta_1, z_{-i})) - c(y_i(\theta_1, z_{-i}), \theta_1)] = 0.$$

By IC-type(out), we have

$$E_{-i}[\sum_x p_i(x,\hat{z}_i,z_{-i})f(x|y_i'(\theta_1,\hat{z}_i,z_{-i}),y_{-i}(\hat{z}_i,z_{-i}))-c(y_i'(\theta_1,\hat{z}_i,z_{-i}),\theta_1)]\le 0,\ \ \forall y_i'(\theta_1,\hat{z}_i,z_{-i}),\ \hat{z}_i\neq\theta_1.$$

Thus,

$$E_{-i}[\sum_x p_i(x,\hat{z}_i,z_{-i})f(x|0,y_{-i}(\hat{z}_i,z_{-i}))-c(0,\theta_1)]\le 0,\ \ \forall\hat{z}_i\neq\theta_1.$$

Since $p_i(x,\hat{z}_i,z_{-i})\ge 0\ \forall x,\hat{z}_i,z_{-i}$ due to LL(out), we have

$$\sum_x p_i(x,\hat{z}_i,z_{-i})f(x|0,y_{-i}(\hat{z}_i,z_{-i}))-c(0,\theta_1)\ge 0,\ \ \forall\hat{z}_i\neq\theta_1,z_{-i}.$$

Thus,

$$\sum_x p_i(x,\hat{z}_i,z_{-i})f(x|0,y_{-i}(\hat{z}_i,z_{-i}))-c(0,\theta_1)=0,\ \ \forall\hat{z}_i\neq\theta_1,z_{-i},$$

implying $p_i(x,\hat{z}_i,z_{-i})=0$, $\forall x,\hat{z}_i,z_{-i}$, where $\hat{z}_i\neq\theta_1$.

For agent i whose true type is $\hat{z}_i$, given that $p_i(x,\hat{z}_i,z_{-i})=0$, $\forall x,\hat{z}_i,z_{-i}$, his best contribution level is 0, and thus, his interim and ex-ante expected payoffs are 0. Thus, if the lowest type of agent i earns 0 information rent, every type of him earns 0 information rent. **Q.E.D.**

Proof of Lemma 2. Take a feasible punishing contract $\{\breve{p}_i(y_i,\hat{z}_i,\hat{z}_{-i}),y_i^c(\hat{z}_i,\hat{z}_{-i})\}$. Suppose there exists an agent i with true type $\tilde{z}_i$ such that all of his IC-type*(con) and PC(con) constraints are non-binding, i.e.,

$$E_{-i}\bar{\pi}_i(\tilde{z}_i,\tilde{z}_i,y_i^c(\tilde{z}_i,z_{-i}),z_{-i})>E_{-i}\bar{\pi}_i(\tilde{z}_i,\hat{z}_i,\hat{y}_i^{con}(\tilde{z}_i,\hat{z}_i,z_{-i}),z_{-i}),\ \ \forall\hat{z}_i\neq\tilde{z}_i,\ \text{and}$$
$$E_{-i}\bar{\pi}_i(\tilde{z}_i,\tilde{z}_i,y_i^c(\tilde{z}_i,z_{-i}),z_{-i})>0,$$

where $\hat{y}_i^{con}(\tilde{z}_i,\hat{z}_i,z_{-i})\in\hat{Y}_i^{con}(\tilde{z}_i,\hat{z}_i,z_{-i})$. Thus, there exists some type profile of other agents such that agent i's interim payoff is strictly positive. Pick one such type profile, say $\check{z}_{-i}$, so $\bar{\pi}_i(\tilde{z}_i,\tilde{z}_i,y_i^c(\tilde{z}_i,\check{z}_{-i}),\check{z}_{-i})>0$. This also implies $\breve{p}_i(y_i^c(\tilde{z}_i,\check{z}_{-i}),\tilde{z}_i,\check{z}_{-i})>0$.

Now, define

$$\delta\equiv\min_{\hat{z}_i\neq\tilde{z}_i}\{E_{-i}\bar{\pi}_i(\tilde{z}_i,\tilde{z}_i,y_i^c(\tilde{z}_i,z_{-i}),z_{-i})-E_{-i}\bar{\pi}_i(\tilde{z}_i,\hat{z}_i,\hat{y}_i^{con}(\tilde{z}_i,\hat{z}_i,z_{-i}),z_{-i})\}>0,$$

which is the smallest difference between agent i's expected equilibrium payoff and expected deviation payoffs obtained among all non-truthfully reported types, and

$$\epsilon\equiv\frac{1}{2}\min\{\delta,\bar{\pi}_i(\tilde{z}_i,\tilde{z}_i,y_i^c(\tilde{z}_i,\check{z}_{-i}),\check{z}_{-i}),E_{-i}\bar{\pi}_i(\tilde{z}_i,\tilde{z}_i,y_i^c(\tilde{z}_i,z_{-i}),z_{-i})\}>0.$$

Consider the following payment rule:

$$\breve{p}_i'(y_i,\hat{z}_i,\hat{z}_{-i})=\begin{cases}\breve{p}_i(y_i^c(\tilde{z}_i,\check{z}_{-i}),\hat{z}_i,\hat{z}_{-i})-\epsilon, & \text{if } (\hat{z}_i,\hat{z}_{-i})=(\tilde{z}_i,\check{z}_{-i}) \text{ and } y_i=y_i^c(\tilde{z}_i,\check{z}_{-i})\\ 0, & \text{if } (\hat{z}_i,\hat{z}_{-i})=(\tilde{z}_i,\check{z}_{-i}) \text{ and } y_i\neq y_i^c(\tilde{z}_i,\check{z}_{-i})\\ \breve{p}_i(y_i,\hat{z}_i,\hat{z}_{-i}), & \text{otherwise,}\end{cases}$$

and $\breve{p}'_j(y_j, \hat{z}_j, \hat{z}_{-j}) = \breve{p}_j(y_j, \hat{z}_j, \hat{z}_{-j})$ for $j \neq i$.

Under this new contract, only the payment to agent i is reduced by a fixed amount ϵ if the reported profile is $(\tilde{z}_i, \check{z}_{-i})$ and agent i chooses $y_i = y_i^c(\tilde{z}_i, \check{z}_{-i})$. Now, check that under the new contract, LL(con) is still satisfied, i.e., $\breve{p}'_i(y_i, \hat{z}_i, \hat{z}_{-i}) \geq 0$. Since $\breve{p}_i(y_i, \hat{z}_i, \hat{z}_{-i})$ is feasible by assumption, $\breve{p}_i(y_i, \hat{z}_i, \hat{z}_{-i}) \geq 0$. Also,

$$
\begin{aligned}
& \breve{p}_i(y_i^c(\tilde{z}_i, \check{z}_{-i}), \tilde{z}_i, \check{z}_{-i}) - \epsilon \\
= \; & \breve{p}_i(y_i^c(\tilde{z}_i, \check{z}_{-i}), \tilde{z}_i, \check{z}_{-i}) - \frac{1}{2}\min\{\delta, \bar{\pi}_i(\tilde{z}_i, \tilde{z}_i, y_i^c(\tilde{z}_i, \check{z}_{-i}), \check{z}_{-i}), E_{-i}\bar{\pi}_i(\tilde{z}_i, \tilde{z}_i, y_i^c(\tilde{z}_i, z_{-i}), z_{-i})\} \\
\geq \; & \breve{p}_i(y_i^c(\tilde{z}_i, \check{z}_{-i}), \tilde{z}_i, \check{z}_{-i}) - \frac{1}{2}\bar{\pi}_i(\tilde{z}_i, \tilde{z}_i, y_i^c(\tilde{z}_i, \check{z}_{-i}), \check{z}_{-i}) \\
= \; & \breve{p}_i(y_i^c(\tilde{z}_i, \check{z}_{-i}), \tilde{z}_i, \check{z}_{-i}) - \frac{1}{2}[\breve{p}_i(y_i^c(\tilde{z}_i, \check{z}_{-i}), \tilde{z}_i, \check{z}_{-i}) - c(y_i^c(\tilde{z}_i, \check{z}_{-i}), \tilde{z}_i)] \\
= \; & \frac{1}{2}[\breve{p}_i(y_i^c(\tilde{z}_i, \check{z}_{-i}), \tilde{z}_i, \check{z}_{-i}) + c(y_i^c(\tilde{z}_i, \check{z}_{-i}), \tilde{z}_i)] \qquad \text{(A.1)} \\
> \; & 0.
\end{aligned}
$$

Thus, the new contract $\{\breve{p}'_i(y_i, \hat{z}_i, \hat{z}_{-i}), y_i^c(\hat{z}_i, \hat{z}_{-i})\}$ satisfies LL(con).

Next, check that under this new contract, the equilibrium contributions for each type of the agent are still the same as that under the original contract $\{\breve{p}_i(y_i, \hat{z}_i, \hat{z}_{-i}), y_i^c(\hat{z}_i, \hat{z}_{-i})\}$. Under $\breve{p}'_i(y_i, \hat{z}_i, \hat{z}_{-i})$, if agent i with true type $\tilde{z}_i$ contributes $y_i^c(\tilde{z}_i, \check{z}_{-i})$, his interim payoff is

$$
\begin{aligned}
& \bar{\pi}'_i(\tilde{z}_i, \tilde{z}_i, y_i^c(\tilde{z}_i, \check{z}_{-i}), \check{z}_{-i}) \\
= & \breve{p}_i(y_i^c(\tilde{z}_i, \check{z}_{-i}), \tilde{z}_i, \check{z}_{-i}) - \epsilon - c(y_i^c(\tilde{z}_i, \check{z}_{-i}), \tilde{z}_i) \\
\geq & \frac{1}{2}[\breve{p}_i(y_i^c(\tilde{z}_i, \check{z}_{-i}), \tilde{z}_i, \check{z}_{-i}) - c(y_i^c(\tilde{z}_i, \check{z}_{-i}), \tilde{z}_i)] \qquad \text{(by (A.1))} \\
= & \frac{1}{2}\bar{\pi}_i(\tilde{z}_i, \tilde{z}_i, y_i^c(\tilde{z}_i, \check{z}_{-i}), \check{z}_{-i}) > 0,
\end{aligned}
$$

which is higher than his payoff from any other contribution choice. Thus, his optimal contribution level is still $y_i^c(\tilde{z}_i, \check{z}_{-i})$. As incentives for other types are not changed, the equilibrium contributions remain the same for all types.

In addition, we can easily check that PC(con) and IC-type(con) constraints for agent i if his true type is $\tilde{z}_i$ are still satisfied, as by construction, the reduction in payment is not big enough to alter his truth-telling and participation incentives. If agent i's true type is not $\tilde{z}_i$, the previous contract $\breve{p}_i(y_i, \hat{z}_i, \hat{z}_{-i})$ ensures that he has no incentive to report as $\tilde{z}_i$. Now, in the new contract $\breve{p}'_i(y_i, \hat{z}_i, \hat{z}_{-i})$, the deviation ex-ante expected payoff becomes even smaller if he reports as type $\tilde{z}_i$, which strengthens the validity of his IC-type(con) constraints.

Therefore, the new contract $\{\breve{p}'_i(y_i, \hat{z}_i, \hat{z}_{-i}), y_i^c(\hat{z}_i, \hat{z}_{-i})\}$ is feasible and induces same amount of contributions with less expected cost. Thus, the principal's ex-ante expected profit will be higher, given our assumption in the paper that each type profile occurs with positive probability (and thus the occurrence of $(\tilde{z}_i, \check{z}_{-i})$ is a non-negligible event). **Q.E.D.**

Proof of Proposition 4. We prove by contradiction. Suppose for a given optimal punishing contract $\{\breve{p}_i(y_i, z_i, z_{-i}), y_i^c(z_i, z_{-i})\}$, the lowest type of agent i earns strictly positive information rent. Thus, every type of agent i should earn strictly positive ex-ante expected payoff, since the higher type can always mimic as the lowest type, choose the target contribution for the lowest type at a weakly lower cost, and thus, earn a weakly higher payoff than the lowest type's equilibrium payoff.

Note that for the interim payoff, the following inequalities are always satisfied:

$$\bar{\pi}_i(z_i, z_i, y_i^c(z_i, z_{-i}), z_{-i}) \geq \bar{\pi}_i(\tilde{z}_i, z_i, \hat{y}_i^{con}(\tilde{z}_i, z_i, z_{-i}), z_{-i}) \quad \forall i, z_{-i}, \text{ if } z_i > \tilde{z}_i; \tag{A.2}$$

$$\bar{\pi}_i(z_i, \tilde{z}_i, \hat{y}_i^{con}(z_i, \tilde{z}_i, z_{-i}), z_{-i}) \geq \bar{\pi}_i(\tilde{z}_i, \tilde{z}_i, y_i^c(\tilde{z}_i, z_{-i}), z_{-i}) \quad \forall i, z_{-i}, \text{ if } z_i > \tilde{z}_i; \tag{A.3}$$

and $$\bar{\pi}_i(z_i, \hat{z}_i, \hat{y}_i^{con}(z_i, \hat{z}_i, z_{-i}), z_{-i}) \geq \bar{\pi}_i(\tilde{z}_i, \hat{z}_i, \hat{y}_i^{con}(\tilde{z}_i, \hat{z}_i, z_{-i}), z_{-i}) \quad \forall i, z_{-i}, \ \hat{z}_i \neq z_i, \tilde{z}_i, \text{ if } z_i > \tilde{z}_i. \tag{A.4}$$

The first inequality says for each agent i, given any particular type profile of other agents, the higher type of him can always obtain a weakly higher interim equilibrium payoff than the interim deviation payoff of those lower type who wants to mimic him. For example, $\bar{\pi}_i(\theta_2, \theta_2, y_i^c(\theta_1, z_{-i}), z_{-i}) \geq \bar{\pi}_i(\theta_1, \theta_2, \hat{y}_i^{con}(\theta_1, \theta_2, z_{-i}), z_{-i})$. The second inequality says for each agent i, given any particular type profile of other agents, when the higher type mimics a lower type, the higher type can always obtain a weakly higher interim deviation payoff than the interim equilibrium payoff of the lower type. For example, $\bar{\pi}_i(\theta_2, \theta_1, \hat{y}_i^{con}(\theta_2, \theta_1, z_{-i}), z_{-i}) \geq \bar{\pi}_i(\theta_1, \theta_1, y_i^c(\theta_1, z_{-i}), z_{-i})$. The last inequality says for each agent i, given any particular type profile of other agents, the interim deviation payoff of higher type is always weakly higher than that of the lower type if they mimic as some other common type. For example, $\bar{\pi}_i(\theta_2, \theta_3, \hat{y}_i^{con}(\theta_2, \theta_3, z_{-i}), z_{-i}) \geq \bar{\pi}_i(\theta_1, \theta_3, \hat{y}_i^{con}(\theta_1, \theta_3, z_{-i}), z_{-i})$. These three inequalities are due to the monotonicity of the cost function in type.

Since the contract is optimal and for each type z_i his PC(con) is non-binding (as noted at the start of this proof), by Lemma 2, at least one of his IC-type*(con) must be binding. This implies at least one of z_i type's ex-ante deviation payoff is positive. Denote R to be the smallest positive ex-ante deviation payoffs across all type pairs $(z_i, \hat{z}_i)$ (where z_i is the true type and $\hat{z}_i \neq z_i$ is the reported type). That is,

$$R \equiv \min_{z_i, \hat{z}_i} \left\{ E_{-i}\bar{\pi}_i(z_i, \hat{z}_i, \hat{y}_i^{con}(z_i, \hat{z}_i, z_{-i}), z_{-i}) \, : \, E_{-i}\bar{\pi}_i(z_i, \hat{z}_i, \hat{y}_i^{con}(z_i, \hat{z}_i, z_{-i}), z_{-i}) > 0 \right\}.$$

For any type $\hat{z}_i$, we know that when some other type z_i reports as type $\hat{z}_i$, either

$$E_{-i}\bar{\pi}_i(z_i, \hat{z}_i, \hat{y}_i^{con}(z_i, \hat{z}_i, z_{-i}), z_{-i}) \geq R,$$

or

$$E_{-i}\bar{\pi}_i(z_i, \hat{z}_i, \hat{y}_i^{con}(z_i, \hat{z}_i, z_{-i}), z_{-i}) = 0.$$

The rest of the proof is divided into two steps.

Step 1. Choose a particular type $\hat{z}_i'$ for which there exists some z_i such that

$$E_{-i}\bar{\pi}_i(z_i, \hat{z}_i', \hat{y}_i^{con}(z_i, \hat{z}_i', z_{-i}), z_{-i}) \geq R.$$

Let

$$\bar{z}_i = \min\left\{z_i \ : \ E_{-i}\bar{\pi}_i(z_i, \hat{z}_i', \hat{y}_i^{con}(z_i, \hat{z}_i', z_{-i}), z_{-i}) \geq R\right\}.$$

Thus, for all types $z_i < \bar{z}_i$,

$$E_{-i}\bar{\pi}_i(z_i, \hat{z}_i', \hat{y}_i^{con}(z_i, \hat{z}_i', z_{-i}), z_{-i}) = 0,$$

and for all types $z_i' > \bar{z}_i$, we know by (A.4) that

$$E_{-i}\bar{\pi}_i(z_i', \hat{z}_i', \hat{y}_i^{con}(z_i', \hat{z}_i', z_{-i}), z_{-i}) \geq R.$$

Consider the following two cases:

Case (i) $\bar{z}_i < \hat{z}_i'$.

Given that $E_{-i}\bar{\pi}_i(\bar{z}_i, \hat{z}_i', \hat{y}_i^{con}(\bar{z}_i, \hat{z}_i', z_{-i}), z_{-i}) \geq R$, identify all z_{-i} profiles such that $\bar{\pi}_i(\bar{z}_i, \hat{z}_i', y_i^c(\hat{z}_i', z_{-i}), z_{-i}) > 0$, which also implies $\breve{p}_i(y_i^c(\hat{z}_i', z_{-i}), \hat{z}_i', z_{-i}) > 0$. The expected sum of the above interim deviation payoffs must be at least R. Now reduce the above payments $\breve{p}_i(y_i^c(\hat{z}_i', z_{-i}), \hat{z}_i', z_{-i})$ such that the expected reductions in the ex-ante deviation payoff of agent $\bar{z}_i$ is R, and yet his optimal contribution response following misreporting as $\hat{z}_i'$ remains the same as before the reduction.

Since $\hat{z}_i' > \bar{z}_i$, by (A.2), we know

$$\bar{\pi}_i(\hat{z}_i', \hat{z}_i', y_i^c(\hat{z}_i', z_{-i}), z_{-i}) \geq \bar{\pi}_i(\bar{z}_i, \hat{z}_i', y_i^c(\hat{z}_i', z_{-i}), z_{-i}), \quad \forall z_{-i}. \tag{A.5}$$

Thus, the expected reduction in the ex-ante equilibrium payoff for type $\hat{z}_i'$ agent is also R.[20] Before the reduction, we know $E_{-i}\bar{\pi}_i(\hat{z}_i', \hat{z}_i', y_i^c(\hat{z}_i', z_{-i}), z_{-i}) \geq R$, so the ex-ante expected payoff for type $\hat{z}_i'$ agent is still non-negative after the reduction, i.e., PC(con) is still satisfied.

Case (ii) $\bar{z}_i > \hat{z}_i'$.

Given that $E_{-i}\bar{\pi}_i(\hat{z}_i', \hat{z}_i', y_i^c(\hat{z}_i', z_{-i}), z_{-i}) \geq R$ (due to the fact that R is the smallest positive deviation payoff), identify all z_{-i} profiles such that $\bar{\pi}_i(\hat{z}_i', \hat{z}_i', y_i^c(\hat{z}_i', z_{-i}), z_{-i}) > 0$, which also implies $\breve{p}_i(y_i^c(\hat{z}_i', z_{-i}), \hat{z}_i', z_{-i}) > 0$. The expected sum of the above interim payoffs must be at least R. Now reduce the above payments $\breve{p}_i(y_i^c(\hat{z}_i', z_{-i}), \hat{z}_i', z_{-i})$ such that the expected reductions in the ex-ante equilibrium payoff of agent $\hat{z}_i'$ is R, and yet his optimal contribution remains the same as before the reduction.

Since $\bar{z}_i > \hat{z}_i'$, by (A.3), we know $\bar{\pi}_i(\bar{z}_i, \hat{z}_i', y_i^c(\hat{z}_i', z_{-i}), z_{-i}) \geq \bar{\pi}_i(\hat{z}_i', \hat{z}_i', y_i^c(\hat{z}_i, z_{-i}), z_{-i}) \ \forall z_{-i}$. Thus, the expected reduction in the ex-ante deviation payoff for type $\bar{z}_i$ agent when reporting as type $\hat{z}_i'$ is also R. $\|$

For all types $z_i < \bar{z}_i$, initially $E_{-i}\bar{\pi}_i(z_i, \hat{z}_i', \hat{y}_i^{con}(z_i, \hat{z}_i', z_{-i}), z_{-i}) = 0$, and after the reduction in payment, those ex-ante deviation payoff of agent z_i when he reports as $\hat{z}_i'$ is still 0.

For all types $z_i' > \bar{z}_i$, initially $E_{-i}\bar{\pi}_i(z_i', \hat{z}_i', \hat{y}_i^{con}(z_i', \hat{z}_i', z_{-i}), z_{-i}) \geq R$. By (A.4), we know $\bar{\pi}_i(z_i', \hat{z}_i', y_i^c(\hat{z}_i', z_{-i}), z_{-i}) \geq \bar{\pi}_i(\bar{z}_i, \hat{z}_i', y_i^c(\hat{z}_i', z_{-i}), z_{-i}) \ \forall z_{-i}$. Thus, the expected reduction in the deviation payoff of agent z_i' when he reports as $\hat{z}_i'$ is R.

[20] Note that the reductions in type $\bar{z}_i$'s deviation payment that were carried out earlier for $(\hat{z}_i', z_{-i})$ reported profiles, the same reductions will happen when the true type is $\hat{z}_i'$ and the reported profiles are the same $(\hat{z}_i', z_{-i})$. That is, the same reductions apply to both sides of (A.5).

Step 2. Repeat such reductions as in Step 1 for each possible type $\hat{z}_i''$ for which there exists z_{-i} such that

$$E_{-i}\bar{\pi}_i(z_i, \hat{z}_i'', \hat{y}_i^{con}(z_i, \hat{z}_i'', z_{-i}), z_{-i}) \geq R.$$

Thus, we know that each such $\hat{z}_i''$ type's ex-ante equilibrium payoff is reduced by R, and the payoff is still non-negative. Also, for the ex-ante deviation payoff that is initially 0, it is still 0 after the reduction in payment; for the ex-ante deviation payoff that is initially positive, it is also reduced by R after the reduction in payment. Thus, every type's IC-type*(con) is still satisfied.

We have thus shown the existence of a feasible punishing contract which generates a higher ex-ante expected profit for the principal (by maintaining the same contributions level with a lower payment), contradicting that the original contract $\{\breve{p}_i(y_i, z_i, z_{-i}), y_i^c(z_i, z_{-i})\}$ is optimal. **Q.E.D.**

■ **The Proof of the main Claim in Vander Veen (1995).** The purpose of this note is to point out that while Vander Veen's argument that risk aversion should prompt the principal to prefer to monitor individual contributions (over total team output) stands to economic reason, the author's formal proof has a serious gap. We do not intend to fix this gap as it is not integral to our main objective. We point it out here as we fail to see if and how Vander Veen's treatment of risk aversion could shed any light on the optimal monitoring choice question when agents' risk aversion is replaced by limited liability constraints.

The reason one may want to make a connection between risk aversion and limited liability is the following: In some sense, limited liability could be considered as a specific instance of agent risk aversion with the utility approaching negative infinity if their payments/rewards become too low. But then not only one has to modify the linear utility functions in our formulation (linear in monetary rewards) by concave segments near zero or negative rewards, one also has to ensure that this *local concavity* is enough to guarantee the dominance of individual contributions monitoring over output monitoring. Below we identify the gap in Vander Veen's proof.

Let us start with the statement of Vander Veen's claim (stated exactly as it appeared in the article):

Claim. *Let $p_i(x(\mathbf{y}), \hat{\mathbf{z}})$ be the McAfee and McMillan payment function which is nonconstant. If the agents are risk averse, and total output is an insufficient statistic for the individual contributions, then there exists a new sharing rule with monitoring of individual contributions which is Pareto improving.*

The benchmark of McAfee and McMillan's payment function $p_i(x(\mathbf{y}), \hat{\mathbf{z}})$ cannot be accurate. Instead, it should be the optimal contract under output monitoring when agents have risk aversion. Let us therefore assume that the $p_i(x(\mathbf{y}), \hat{\mathbf{z}})$ function in the above claim is the **optimal contract under risk aversion**.

Let us now return to our main concern. For his domination argument, the author defines a new payment rule as follows (see p. 1055 of Vander Veen's article):

$$\tilde{p}(\mathbf{y}, \hat{\mathbf{z}}) = p(\mathbf{y}, \hat{\mathbf{z}}) + I_{11}(\mathbf{y})ds_{11} + I_{12}(\mathbf{y})ds_{12} + I_2(\mathbf{y})ds_2$$

$$\textit{and} \quad \tilde{p}_{\hat{z}} = k \quad \textit{where } k \in \mathcal{R} \textit{ is a constant},$$

$$I_{i,j} = 1 \quad \textit{if } \mathbf{y} \in \{\mathbf{y}_{11}, \mathbf{y}_{12}\},$$

$$0 \quad \textit{otherwise}$$

and ds_{11}, ds_{12}, *and* ds_2 *are chosen appropriately.*

The author's restriction that $\tilde{p}_{\hat{z}} = k$ is saying that the newly constructed payment function is linear in agent types ($\hat{\mathbf{z}}$ denoting the type profile). Later on in his proof the author uses the property that $p_{\hat{z}} = k \quad \forall X_j$ (see p. 1056), which suggests that the author will have to restrict the payment function $p(x(\mathbf{y}), \hat{\mathbf{z}})$ to be linear in types.[21] But then such linearity has to be the property of the optimal payment function under risk aversion that is being dominated. Without establishing such property explicitly, it is unclear how the author can simply impose this restriction. This leaves one to doubt how the author's argument can be taken to apply to the original optimal payment function under risk aversion, which the author never derives explicitly.

Besides the above concern, the other reason we consider Vander Veen's risk aversion argument cannot apply to our setting is, he seems to be using global risk aversion. ■

References

[1] Bertrand, M. and Mullainathan, S., "Are CEOs Rewarded for Luck? The Ones without Principals Are." *Quarterly Journal of Economics* 116 (2001), 901-932.

[2] Holmström, B., "Moral Hazard in Teams." *Bell Journal of Economics* 13 (1982), 324-340.

[3] Innes, R.D., "Limited Liability and Incentives Contracting with Ex-ante Action Choices." *Journal of Economic Theory* 52 (1990), 45-67.

[4] Khalil, F. and Lawarrée, J., "Input versus Output Monitoring: Who is the Residual Claimant?" *Journal of Economic Theory* 66 (1995), 139-157.

[5] Lawarrée, J. and Audenrode, M. Van., "Optimal Contract, Imperfect Output Observation, and Limited Liability." *Journal of Economic Theory* 71 (1996), 514-531.

[6] Lewis, T. and Sappington, D., "Motivating Wealth-Constrained Actors." *American Economic Review* 90 (2000), 944-960.

[7] McAfee, R.P. and McMillan, J., "Optimal Contracts for Teams." *International Economic Review* 32 (1991), 561-577.

[8] Ollier, S. and Thomas, L., "Ex Post Participation Constraint in a Principal-Agent Model with Adverse Selection and Moral Hazard." *Journal of Economic Theory* 148 (2013), 2383-2403.

[21] Note that the author had started out with the linearity restriction on $\tilde{p}(\mathbf{y}, \hat{\mathbf{z}})$, then switched to $p(x(\mathbf{y}), \hat{\mathbf{z}})$.

[9] Sappington, D., "Limited Liability Contracts between Principal and Agent." *Journal of Economic Theory* 29 (1983), 1-21.

[10] Vander Veen, T.D., "Optimal Contacts for Teams: A Note on the Results of McAfee and McMillan." *International Economic Review* 36 (1995), 1051-1056.