

## How to Restore Higher-Powered Incentives in Multitask Agencies

Bernard Sinclair-Desgagné  
École Polytechnique and CIRANO, Montréal

In multiple-task agency setups it is commonly accepted that wage incentives must be weaker when the agent's performance on some of the activities is difficult to measure. This article shows that stronger incentives can be restored through a scheme of selective audits in which the appraisal of less tangible activities is contingent on observing high performance levels in the more visible tasks. This scheme would make the efforts expended on the various tasks complementary rather than substitutes in the agent's utility function. It is optimal under plausible assumptions concerning the monitoring technology (separability of the multivariate likelihood function) and the agent's risk behavior (absolute prudence larger than three times absolute risk aversion).

Then Étienne began to read him the announcement. It was a notice from the Company to the miners of all the pits, informing them that *in consequence of the lack of care bestowed on the timbering*, the Company had resolved to apply a new method of payment for the extraction of coal. Henceforward *the price of the tub of coal extracted would be lowered*, from fifty centimes to forty, and the Company itself would pay for the timber.

[based on Émile Zola, *Germinal*, Chapter III]

### 1. Introduction

Employment contracts often make compensation vary relatively little across performance levels. This fact is perceived to be so typical that common wisdom

---

This article was previously circulated under the title, "Restoring Higher-Powered Incentives through Audits." I am grateful to Marcel Boyer, Avinash Dixit, Landis Gabel, Michael Huberman, Thomas Jeitschko, Dilip Mookherjee, Elisabeth Paté-Cornell, Lambros Pechlivanos, Rafael Rob, Michel Roland, Kathleen Segerson, David Soskice, Jean Tirole, Nicolas Treich, David Ulph, a coeditor, and an anonymous referee for useful comments. I also wish to thank the Wissenschaftszentrum-Berlin, where part of this work was carried out. This research was made possible by a strategic research grant from the Social Sciences and Humanities Research Council of Canada.

frequently opposes the so-called “low-powered incentives” of firms to the more responsive rewards, or “high-powered incentives,” which characterize the market.

The economic literature provides several explanations for the prevalence of lower-powered incentives in nonmarket organizations. Early agency theory (Ross, 1973; Mirrlees, 1976; Holmström, 1979; Shavell, 1979) emphasizes the trade-off between risk sharing and incentives: the latter implies that agents must foresee some punishments and rewards in order to provide effort, but the former requires that agents, who are usually more risk averse than the firm, be somewhat sheltered against random fluctuations of the outcome. In his more recent book, Williamson (1985) argues that incentives are milder in firms for two additional reasons: first, measurement problems might preclude credible commitments to pure performance-based compensation; and second, marketlike sharper incentives would lead to distortions in asset utilization and the innovation process. The latter rationale suggests that incentives must generally be weaker when effort and output on a given job have many dimensions. Such an argument underlies the above *Germinal* story: the company wants the miners to make a better allocation of effort between safety (timbering) and production (coal extraction), and to accomplish this it reduces the piece rate on the coal extracted. The multiple-task rationale was explored and first formalized by Holmström and Milgrom (1991). In a recent article that builds on this work, Dixit (1997) shows that the interaction among several stakeholders—a feature deemed to be frequent in public bureaucracies—results in a further loss of the power of incentives.

Among these explanations, the existence of multidimensional efforts and outcomes stands out so far as the most compelling justification for the adoption of low-powered incentives in nonmarket organizations. Based on their previous work (Holmström and Milgrom, 1987), Holmström and Milgrom (1991) have developed a tractable multiple-task principal-agent model where the agent’s effort and performance possess several dimensions. Their model predicts that<sup>1</sup>

(i) “(. . .) when there are multiple tasks, incentive pay serves not only to allocate risks and to motivate hard work, it also serves to direct the allocation of the agents’ attention *among* their various duties.”

(ii) “(. . .) the desirability of providing incentives for any one activity decreases with the difficulty of measuring performance in any other activities that make competing demands on the agent’s time and attention.”

These propositions have found empirical support in various contexts, such as salesforce management (Anderson and Schmittlein, 1984), franchising (Krueger, 1991), and retailing contracts (Slade, 1996).

---

1. The model is richer than the following two statements seem to suggest. It also allows for a treatment of ownership issues and job design.

Holmström and Milgrom (1991) also deal with the normative issue of how to increase the power of incentives in a multitask agency setting. They suggest two methods. One is to reduce the dilution of effort by putting explicit restrictions on some activities. This is an obvious purpose of the procedures and constraints large firms and bureaucracies often impose on employees in order to regulate their work.<sup>2</sup> There seems to be real limitations to such a method, however, as persistent information asymmetry between managers and workers concerning the latter's work habits generates frictions and makes it increasingly costly to apply (see, for instance, Miller, 1992:112–13). Another possible way is to regroup tasks and define jobs in such a way that each agent's job includes tasks that differ as little as possible in their measurement characteristics; the intensity of incentives for a given job should then increase with the ease of measuring performance on that job. This second method also seems impractical in most concrete situations, because of the imperatives of technology (Lazear, 1995), or because it usually requires the resolution of a highly complex combinatorial problem (Karp, 1972).

These two methods are based on altering the agent's opportunity cost of effort on the various tasks. This article now takes the other route, which consists of changing the way agents are compensated. The alternative control and reward scheme that we propose can be described as follows.

- Consider an agent whose job includes two tasks, A and B.
- Let task A performance be routinely monitored, while performance on task B can be audited.
- The principal commits to making an audit only when task A performance is assessed to be high.
- The agent's expected utility is higher when an audit takes place than when no audit occurs.
- However, if an audit yields a bad assessment of performance on task B, then the agent's ex post wage is inferior to what it would have been if no audit had taken place.

The intuitive reason why this scheme might restore the power of incentives is straightforward. Clearly an agent should want to be audited under the above scheme. He would then be led to spend more effort on task A (ignoring for the moment his cost of effort) in order to increase the likelihood of showing high performance on this task. But since there is no benefit to being audited if performance on task B is ultimately assessed to be low, he would be led to work harder as well on task B! This means that the respective efforts expended on tasks A and B have now become *complementary*, as far as compensation is

---

2. Casual evidence of the widespread use of this method could again be found in a front-page article of the September 23, 1997, issue of the *International Herald Tribune*. The article ["Goofing Off at Work: First You Log On"] reported that "Employers, both in the federal government and in the private sector, are cracking down on the use of computer games, personal e-mail and recreational Internet surfing, which they see as undermining the productivity that the PC was supposed to bring to the world of work."

concerned. If the complementarity is strong enough, it might counterbalance the fact that efforts on tasks A and B are substitutes in the agent's cost function. In this case the conflict between tasks emphasized by Williamson (1985) and by Holmström and Milgrom (1991) would disappear, so the power of incentives could be raised.

This scheme presents several practical advantages over the previous approaches. First, it might be quite costly to amend downward an already functioning incentive scheme. The *Germinal* novel provides a dramatic illustration of this fact, as the lower piece rate meets with the miners' fierce opposition and triggers a bloody strike. Implementing instead the alternative proposed here might have avoided this.<sup>3</sup> With the present scheme, the company could keep the price of the tub of coal unchanged; an inspection of the timbering would occur whenever the amount of coal extracted by a given team of miners gets *above* a certain level; and although some penalties for bad timbering would be imposed ex post, the miners should see ex ante such an inspection as a reward. The miners might then accept to work harder on production and safety *at the same time*, making all parties better off. A second advantage of the actual scheme is that it does not clash with production technology. Consider, for instance, Lazear's (1995:86–87) criticism of job design based on measurement criteria:

... the farm worker who selects the fruit to be picked also does the actual picking. It is conceivable that there are separate spotters, who determine which fruit is to be picked, and pickers, who do the actual picking. But *the tasks are grouped into the one job because separate spotting and picking would necessitate duplication of effort*. Spotting is difficult to monitor because one cannot easily estimate the number of good fruits that the spotter missed. The output of picking is easily observable. The Holmström–Milgrom point suggests therefore that spotting and picking should not be paired, but *technological considerations probably dominate in dictating job structure*. [emphasis added]

With the present scheme, spotting and picking do not need to be assigned to different people. Instead, picking can be monitored and spotting can be audited whenever the quantities picked are large enough; the farm worker gets a higher expected utility when some verification of his spotting takes place, although an indication of sloppiness in that task would leave him with a loss ex post. The two tasks might then become complementary, so high-powered incentives could be put on picking with no undesirable impact on spotting.

The proposed scheme is formalized below in a multitask principal-agent framework. We provide a rigorous yet simple explanation for why contingent assessments of performance on some tasks can create complementarities more

---

3. We are only discussing the economics of the actual story. We would certainly not pretend to be making any suggestion on how to improve the literary quality of the novel.

easily than when there is constant and simultaneous monitoring of performance on all tasks. Unlike Holmström and Milgrom (1991), we do not assume that the agent exhibits constant absolute risk aversion but we a priori consider any type of risk-averse behavior. We find that the agent's effort allocation in this context depends on the relative magnitude of his prudence (as defined in Kimball, 1990) versus his aversion to risk.<sup>4</sup> We show that the above auditing scheme is optimal when the coefficient of absolute prudence is greater than three times the coefficient of absolute risk aversion.

This article is organized as follows. The next section presents the notation, the model, and the main assumptions. Section 3 explains how complementarities between efforts can be created in this setting by contingent assessments of performance on some of the tasks. Section 4 demonstrates the optimality of the proposed auditing scheme under specific conditions concerning the monitoring technology and the agent's risk behavior. Section 5 contains concluding remarks.

## 2. A Multitask Principal-Agent Model

Consider a one-period principal-agent relationship in which the agent must split his effort between two tasks, A and B. Effort on task A is denoted  $a$ , and that on task B is denoted  $b$ . Assessed performance on tasks A and B is summarized by the noisy signals  $i$  and  $j$ , respectively, where  $i = 0, 1, \dots, I$  and  $j = 0, 1, \dots, J$ . Those signals are linked to the agent's efforts through the conditional likelihood function  $g(i, j | a, b)$ . This function is positive and twice continuously differentiable in  $(a, b)$  for all values of  $i, j, a$ , and  $b$ . We also make the following assumption.

*Assumption 1.*  $g$  is separable with respect to efforts; that is  $g(i, j | a, b) = p_i(a) \cdot q_j(b)$ .

The importance of this assumption will be discussed later. Let us simply say for now that it is less restrictive than it might appear. It does not imply, for instance, that nothing can be inferred about  $b$  from observing only the value of  $i$ , provided  $a$  and  $b$  are linked through the agent's cost of effort. It implies, however, that  $j$  and  $i$  are sufficient statistics for  $b$  and  $a$ , respectively (De Groot, 1970).<sup>5</sup>

The principal is risk neutral. She cannot observe the agent's efforts directly. Instead, she routinely gathers signal  $i$ , and she may get the additional assess-

4. Prudence refers to "the propensity to prepare and forearm oneself in the face of uncertainty, in contrast to risk aversion, which is how much one dislikes uncertainty and would turn away from uncertainty if possible" (Kimball, 1990:54). Kimball shows that there is an isomorphism between the theory of risk aversion and that of prudence. Formally, let  $k$  be the agent's current wealth,  $\theta$  a random variable with mean 0, and  $E$  the expectation operator. The equivalent risk premium  $\alpha$  for  $\theta$  and  $k$  is defined as  $EU(k + \theta) = U(k - \alpha)$ , while the equivalent precautionary premium  $\beta$  for  $\theta$  and  $k$  satisfies  $EU'(k + \theta) = U'(k - \beta)$ .

5. As argued elsewhere (Sinclair-Desgagné and Gabel, 1997), for instance, sufficiency of the signal gathered through an audit is an implicit requirement of the increasingly popular ISO norms on quality and environmental management.

ment  $j$  through an audit. Such an audit costs  $K \geq 0$ . Her expected revenue conditional on observing  $i$  and  $j$  is given by  $R(i, j)$ , where  $R(\cdot, \cdot)$  is an increasing function. She can commit to a contract which includes a contingent probability  $m_i$  of making an audit, a wage schedule  $s_i$  in case no audit occurs, and contingent wages  $w_{ij}$  when an audit takes place. For a given contract, noted  $[s_i, w_{ij}, m_i; a, b]$ , her expected net benefit is then

$$V(s_i, w_{ij}, m_i; a, b) = \sum_{i,j} p_i(a)q_j(b) [m_i(R(i, j) - w_{ij} - K) + (1 - m_i)(R(i, j) - s_i)]. \quad (1)$$

The agent is risk averse, with positive, strictly concave and three-time differentiable Von Neumann–Morgenstern utility index  $u(\cdot)$  defined over monetary payoffs. His cost of effort is given by the positive, strictly convex and twice continuously differentiable function  $c(a, b)$ . This function exhibits strict substitutability (supermodularity) in efforts, that is,  $c_{ab} > 0$ .<sup>6</sup> For a given contract, the agent’s expected utility is then

$$U(s_i, w_{ij}, m_i; a, b) = \sum_{i,j} p_i(a)q_j(b) [m_i u(w_{ij}) + (1 - m_i)u(s_i)] - c(a, b). \quad (2)$$

The current contract between the principal and the agent is called optimal if it maximizes the principal’s net benefit under the usual incentive compatibility and participation constraints. Formally, such a contract must be a solution to the problem:

$$\begin{aligned} & \underset{s,w,m;a,b}{\text{maximize}} V(s_i, w_{ij}, m_i; a, b), \quad \text{subject to :} \\ & \text{incentive compatibility: } (a, b) \text{ maximizes } U(s_i, w_{ij}, m_i; a, b) \\ & \text{participation: } U(s_i, w_{ij}, m_i; a, b) \geq U^*. \end{aligned} \quad (3)$$

To solve this problem, we shall replace the incentive compatibility constraint by two inequalities involving the first-order derivatives of  $U(\cdot)$  with respect to  $(a, b)$ . We shall then focus on contracts that solve the “first-order problem”:<sup>7</sup>

$$\begin{aligned} & \underset{s,w,m;a,b}{\text{maximize}} V(s_i, w_{ij}, m_i; a, b), \quad \text{subject to :} \\ & \sum_{i,j} p'_i(a)q_j(b) [m_i u(w_{ij}) + (1 - m_i)u(s_i)] - c_a \geq 0 \\ & \sum_{i,j} p_i(a)q'_j(b) [m_i u(w_{ij}) + (1 - m_i)u(s_i)] - c_b \geq 0 \\ & U(s_i, w_{ij}, m_i; a, b) \geq U^*. \end{aligned} \quad (4)$$

6. Subscripts  $a$  and  $b$  denote partial derivatives with respect to  $a$  and  $b$ , respectively.

7. Hence we ignore those contracts where the agent would not put any effort on one of the tasks, that is,  $a$  or  $b = 0$  and  $U_a$  or  $U_b < 0$  at an optimal solution of problem (3). In our point of view, a multitask principal-agent problem that yields this solution only would be ill-defined.

This method is called the first-order approach. It is guaranteed to work in this context under the following assumptions (see Sinclair-Desgagné, 1994).<sup>8</sup>

*Assumption 2.*  $p'_i(a)/p_i(a)$  and  $q'_j(b)/q_j(b)$  are nondecreasing in  $i$  and  $j$ , respectively.

*Assumption 3.* The matrix  $[G_{ab}(i, j \mid a, b)]$ , where  $G(i, j \mid a, b) = \sum_{k=i}^I \sum_{t=j}^J p_k(a) \cdot q_t(b)$ , is negative semidefinite for all  $i, j, a$ , and  $b$ .

*Assumption 4.* The Lagrange multipliers associated with the two constraints corresponding to the first-order conditions have the same sign.

*Assumption 5.* A solution to problem (4) exists.

Assumption 2 is the familiar monotone likelihood ratio property (MLRP), a property that is commonly assumed in principal-agent analysis. Note that, for one-dimensional distributions only, MLRP implies the stochastic dominance condition (SDC):  $\sum_{k=i}^I p'_k(a)$  and  $\sum_{k=j}^J q'_k(b)$  are nonnegative for every  $i$  and  $j$ , respectively. Assumption 3 generalizes (and *weakens*) another common assumption: the convexity of the distribution function condition (CDFC). Assumption 4 is necessary here because the first-order conditions involve *two* inequalities. Finally, Assumption 5, together with the other assumptions, ensures that there is an optimal contract.

### 3. Selective Audits and the Creation of Task Complementarity

In a multitask principal-agent context incentives are set by the principal in order to regulate the agent's effort supply on each task. What matters here is not only the respective levels of these supplies, but also their interaction. The latter is captured (locally) by the cross partial derivative  $U_{ab}$  of the agent's expected utility defined in Equation (2). When  $U_{ab} < 0$ , inducing greater effort on one task would cause substitution away from the other task, so raising the strength of incentives on task A by a small amount would decrease the average performance on task B. This corresponds to the situation currently emphasized in the literature. When  $U_{ab} > 0$ , on the other hand, efforts are complementary, so the agent would not work harder at the margin on one task without expending more effort on the other task as well. Putting slightly stronger incentives on task A in this case would therefore not be detrimental to task B. Such a win-win situation seems difficult to achieve, however, because the agent's cost of effort  $c(a, b)$  is a strictly supermodular function, which means that increasing effort on one task always raises the marginal cost of effort on the other task.

Under traditional incentive schemes, where there are no contingent audits and both tasks are monitored simultaneously, it is actually too expensive in general for the principal to try to overcome the substitution effect entailed by

8. One needs also that the probability  $m_i$  be nondecreasing in  $i$ , which is the case in the audit scheme derived below.

the cost function. To see why, set  $m_i \equiv 1$  in Equation (2); the agent's expected utility under a contract where task B is always audited is given by

$$U(w_{ij}; a, b) = \sum_{i,j} p_i(a)q_j(b)u(w_{ij}) - c(a, b). \quad (5)$$

The cross-partial derivative of  $U$  with respect to  $a$  and  $b$  can now be written:<sup>9</sup>

$$\begin{aligned} U_{ab}(w_{ij}; a, b) &= \sum_{i,j} p'_i(a)q'_j(b)u(w_{ij}) - c_{ab} \\ &= \sum_{i=1}^I \sum_{j=1}^J [(u(w_{ij}) - u(w_{ij-1})) - (u(w_{i-1j}) - u(w_{i-1j-1}))] \\ &\quad \cdot \sum_{k=j}^J q'_k(b) \sum_{r=1}^I p'_r(a) - c_{ab}. \end{aligned} \quad (6)$$

By Assumption 2 (which implies SDC), a necessary condition for  $U_{ab}$  to be positive is that

$$\text{for some } i, j > 0: u(w_{ij}) - u(w_{ij-1}) > u(w_{i-1j}) - u(w_{i-1j-1}). \quad (7)$$

That is, at some positive values of  $i$  and  $j$  the difference between the wage at performance level  $j$  and the wage at performance level  $j - 1$  should increase with the value of  $i$ . The required size of the increment might actually be significant for an incentive wage that increases sharply in  $i$ , because the utility index  $u$  is a strictly concave function. In general this may not be worthwhile from the principal's viewpoint (see the Appendix).<sup>10</sup> In the constant absolute risk aversion case considered by Holmström and Milgrom (1991), for example, the optimal incentive wage yields nonincreasing differences  $u(w_{ij}) - u(w_{ij-1})$ , which implies that  $U_{ab} \leq -c_{ab} < 0$ .

Consider now the possibility of having contingent audits. Taking the cross-partial derivative of Equation (2) with respect to  $a$  and  $b$ , we have

$$U_{ab}(s_i, w_{ij}, m_i; a, b) = \sum_{i,j} p'_i(a)q'_j(b)m_i [u(w_{ij}) - u(s_i)] - c_{ab}. \quad (8)$$

By Assumption 2, the derivatives  $p'_i(a)$  and  $q'_j(b)$  are negative for low values of  $i$  and  $j$  and positive for high values of  $i$  and  $j$ . The cross-partial derivative  $U_{ab}$  might now be made positive with the following auditing scheme:

- set  $m_i > 0$  only if  $p'_i(a) > 0$ ;
- for all  $i$ , let  $[u(w_{ij}) - u(s_i)]$  have the same sign as  $q'_j(b)$ .

9. The second expression on the right-hand side comes from applying the "convolution" formula used in Sinclair-Desgagné (1994:463–64) twice.

10. Of course, compensation schemes that create complementarity between tasks can always be obtained in models with simultaneous monitoring, at the cost of making specific assumptions. MacDonald and Marx (1998) recently derived such a scheme. Although their model is rather peculiar, their scheme is intuitive and, as expected from expression (7), associates large rewards with success on *all* dimensions.

Under this scheme, audits are triggered by high assessments of performance on task A, and the wage after an audit is undertaken is higher (lower) than the wage when no audit has taken place if performance on task B is found to be good (bad). Whether and when such a scheme turns out to be optimal is the focus of the upcoming section.

#### 4. When Selective Audits Are Optimal

Some features of the above scheme are common to all auditing schemes derived under problem (4) and the previous assumptions. This is the case for the monotonicity of wages in the received signals, and also for the property that incentive wages under an audit have a greater range than wages without an audit. Those features are formally derived in the next proposition.

*Proposition 1.* The wage schedules  $s_i$  and  $w_{ij}$  are nondecreasing in the signals  $i$  and  $j$ . Moreover,  $w_{ij} < s_i$  when  $j$  is low and  $w_{ij} > s_i$  when  $j$  is high.

*Proof.* Let  $\gamma$ ,  $\lambda$ , and  $\mu$  be the Lagrange multipliers associated with the first, second, and third constraints of problem (4), respectively. The first-order condition for the  $s_i$ 's in this problem is

$$(1 - m_i) \sum_j [-p_i q_j + \gamma u'(s_i) p'_i q_j + \lambda u'(s_i) p_i q'_j + \mu u'(s_i) p_i q_j] = 0. \quad (9)$$

The same condition for the  $w_{ij}$ 's is given by

$$m_i [-p_i q_j + \gamma u'(w_{ij}) p'_i q_j + \lambda u'(w_{ij}) p_i q'_j + \mu u'(w_{ij}) p_i q_j] = 0. \quad (10)$$

After a little bit of algebra, these two conditions reduce, respectively, to

$$\forall i: \frac{1}{u'(s_i)} = \mu + \gamma \frac{p'_i(a)}{p_i(a)} \quad (11)$$

for the  $s_i$ 's, and to

$$\forall i, j: \frac{1}{u'(w_{ij})} = \mu + \gamma \frac{p'_i(a)}{p_i(a)} + \lambda \frac{q'_j(b)}{q_j(b)} \quad (12)$$

for the  $w_{ij}$ 's, without loss of generality. Monotonicity of the wage schedules then follows from Assumption 2, together with the facts that the utility index  $u$  is concave and the Lagrange multipliers are nonnegative.

To prove the second assertion, subtract Equation (12) from Equation (11). This yields

$$\forall i, j: \frac{1}{u'(s_i)} - \frac{1}{u'(w_{ij})} = -\lambda \frac{q'_j(b)}{q_j(b)}. \quad (13)$$

By Assumption 2 (MLRP) and since  $\sum_j q'_j(b) = 0$ , it must be that  $q'_j(b)$  is negative when  $j$  is low and positive when  $j$  is high. The assertion then follows from Equation (13) and the facts that  $u$  is concave and  $\lambda$  is nonnegative. ■

One peculiar feature of the auditing scheme outlined in the introduction, however, is that the agent should *want* to be audited. This might only happen at a significant cost to the principal: the agent being risk averse, the expected wage when an audit takes place but before audit results are known must be strictly higher than the wage received if there had been no audit. It seems therefore that this cost to the principal would usually outweigh the benefits. But as we shall see, there is a plausible range of circumstances where this is not the case.

First note that the agent can protect himself against the additional risk entailed by the possibility of an audit in two ways. One is to try to lower as much as possible the probability that an unfavorable audit conclusion occurs; this behavior is associated with risk aversion. Another mean is to seek and guarantee for oneself the highest expected income; this approach corresponds to the agent's so-called precautionary motives. By offering the agent a wage which is higher on average under an audit than the wage received if no audit occurs, the principal relies on the latter behavior. This strategy might work at a reasonable cost provided the agent's precautionary motives dominate sufficiently his aversion to risk.

To make the latter statement precise, let us recall that the agent's degree of risk aversion can be measured by the coefficient of absolute risk aversion, noted  $r(z) = -u(z)''/u'(z)$ . Let us also introduce the coefficient of absolute prudence, noted  $p(z) = -u'''(z)/u''(z)$ , which can be shown to capture the strength of the agent's precautionary motives (Kimball, 1990). The statement that precautionary motives "dominate" risk aversion in some sense will now be made a formal assumption.

*Assumption 6.*  $\forall z \in \mathbb{R}_+ : p(z) > 3r(z)$ .

All utility indices displaying constant relative risk aversion (CRRA) inferior to  $1/2$ , that is,  $u(z) = z^{1-\rho}$  with  $\rho < 0.5$ , satisfy this assumption.<sup>11</sup> More generally, all utility functions belonging to the so-called hyperbolic absolute risk aversion (HARA) class, that is,

$$u(z) = \frac{1-k}{(2-k)A} [Az + B]^{(2-k)/(1-k)} \quad \text{with } A > 0,$$

fulfill Assumption 6 provided  $k > 3$ .<sup>12</sup> Note, furthermore, that

$$\frac{r'(z)}{r(z)} = \frac{d}{dz} \ln \left( \frac{-u''(z)}{u'(z)} \right) = \frac{u'''(z)}{u''(z)} - \frac{u''(z)}{u'(z)} = r(z) - p(z). \quad (14)$$

Therefore, absolute risk aversion is decreasing—an intuitive assumption that has received strong empirical support (Arrow, 1965:35)—if and only if

11. The coefficient of relative risk aversion is defined as  $\rho = -zu''(z)/u'(z)$ . It measures the changes in the agent's risk attitude when his initial wealth *and* the size of the gamble change together proportionally, by contrast with the coefficient of absolute risk aversion where the size of the gamble is uncorrelated to the agent's initial wealth.

12. However, the CARA (constant absolute risk aversion) utility index used by Holmström and Milgrom (1991) does *not* satisfy Assumption 6, which suggests that working with a general but tractable multitask principal-agent model might be worthwhile.

$p(z) > r(z)$ ; and we have  $p(z) > 3r(z)$  if absolute risk aversion decreases rapidly enough. Consistent with this, Assumption 6 also implies that the Debreu and Koopmans (1982) “concavity index”  $-u''(z)/(u'(z))^2$  is decreasing, and that the marginal propensity to consume out of transitory income declines with the level of wealth (or that the consumption function is strictly concave), a property strongly emphasized in macroeconomics (see Carroll and Kimball, 1996).

We are now ready to show that it is optimal for the principal to offer the agent an expected utility (wage) under an audit which is higher than the corresponding utility (wage) if no audit takes place. The argument of the proof extends that of Dye (1986:345).

*Lemma.* For all  $i$  and  $b$ :  $\sum_j q_j(b)u(w_{ij}) > u(s_i)$ .

*Proof.* By Equations (11) and (12), we have that

$$s_i = \Psi \left( \mu + \gamma \frac{p'_i(a)}{p_i(a)} \right) \quad \text{and} \quad w_{ij} = \Psi \left( \mu + \gamma \frac{p'_i(a)}{p_i(a)} + \lambda \frac{q'_j(b)}{q_j(b)} \right);$$

$$\Psi(\cdot) = \left( \frac{1}{u'(\cdot)} \right)^{-1}.$$

Now note that one consequence of Assumption 5 is that the function  $u(\Psi(\cdot))$  is a convex function. The statement thus follows from Jensen’s inequality. ■

This result is key to show that optimal audits in this case are upper-tailed, that is, that they are triggered by high-performance assessments on task A. Of course, it presupposes that audits are always noisy, even when the agent’s effort on task B is minimal. The next proposition now asserts that, if the probability of occurrence of an audit is positive at performance assessment  $i$  and 0 at  $i'$ , then  $i' < i$ . Hence, when the auditing cost  $K$  is neither too large nor too small, audits only happen at high values of  $i$ . The argument of the proof parallels that of Baiman and Demski (1980).

*Proposition 2.* If  $m_{i'} = 0 < m_i$ , then we must have that  $i' < i$ .

*Proof.* The first-order conditions for the  $m_i$ ’s in problem (4) are given by

$$\left( s_i - K - \sum_j q_j w_{ij} \right) + \left( \sum_j q_j u(w_{ij}) - u(s_i) \right) \left( \mu + \gamma \frac{p'_i}{p_i} \right)$$

$$+ \lambda \sum_j q'_j u(w_{ij}) \geq 0 \quad m_i = 1$$

$$= 0 \quad \text{if} \quad 0 < m_i < 1$$

$$\leq 0 \quad m_i = 0.$$

Substituting Equations (11) and (12), the left-hand expression reduces to

$$s_i - \frac{u(s_i)}{u'(s_i)} + \sum_j q_j \left[ -w_{ij} + \frac{u(w_{ij})}{u'(w_{ij})} \right] - K \tag{15}$$

after some straightforward algebra. Taking the latter’s derivative with respect to  $i$  (doing as if it belonged to a continuum) we get

$$\frac{uu''s'_i}{(u')^2} + \sum_j q_j \left[ -\frac{uu''w'_{:j}}{(u')^2} \right]. \tag{16}$$

By Equation (13), we have that

$$\frac{u''(s_i)s'_i}{(u'(s_i))^2} = \frac{u''(w_{ij})w'_{:j}}{(u'(w_{ij}))^2}. \tag{17}$$

Hence expression (16) is equivalent to

$$-\frac{u''(s_i)s'_i}{(u'(s_i))^2} \left[ -u(s_i) + \sum_j q_j u(w_{ij}) \right]. \tag{18}$$

The latter expression is now positive by the lemma and Proposition 1, which means that expression (15) is increasing in  $i$ . Considering the above first-order conditions, it must then be true that  $m_{i'} = 0 < m_i$  entails  $i' < i$ , as asserted. ■

We shall end this section by discussing the relative importance of the assumptions for the optimality of the current scheme. Assumptions 3, 4, and 5 are technical and are simply meant to guarantee the validity of the first-order approach. Assumption 2 (MLRP) is key for the results that wages are increasing in performance assessments and that wages under an audit have a greater range than those with no audit; but it is not peculiar to the present context. Assumption 1, however, is specific and crucial. The present scheme may not achieve its purpose if no bijection exists between signals and tasks that renders the likelihood function  $g(i, j | a, b)$  separable in both signals and efforts. To illustrate this point briefly, suppose for instance that  $g(i, j | a, b) = p_i(a, b) \cdot q_j(b)$ . Under the suggested auditing scheme, the agent might now care little about task A and put almost all his effort on task B, for this would increase both the probability that a large value of  $i$  triggers an audit and the probability that the audited signal  $j$  is a good one. Finally, Assumption 6 also clearly matters for the scheme’s optimality. Under the reverse assumption for instance,<sup>13</sup> that is,  $\forall z: p(z) < 3r(z)$ , it can be shown that the principal would rather pick a lower-tailed auditing scheme, and that the agent’s expected utility when such an audit occurs would be inferior to the utility achieved otherwise (Sinclair-Desgagné and Gabel, 1997). However, this assumption is only a sufficient condition; although we have been unable to disprove its necessity, it seems plausible that the auditing scheme could remain optimal with a weaker hypothesis.

---

13. Note that the inequalities  $p(z) >$  and  $< 3r(z)$  are *functional* inequalities. Hence they do not cover all cases: it is possible that one given utility function would satisfy each inequality at different points.

## 5. Concluding Remarks

The presence of low-powered incentives in firms and bureaucracies is often justified on the ground that employees of such organizations face multiple tasks, some that can be appraised easily and accurately but others that cannot. In this context, if efforts on each task are substitutes, in the sense that putting more effort on a given task raises the marginal cost of effort applied elsewhere, then current wisdom recommends that strong incentives be avoided because they would bring employees to focus too much on measurable results (number of units sold or produced, short-run benefits, etc.) at the expense of important but less tangible ones (brand reputation, plant and machinery maintenance, long-run financial and technological risks, workers' safety, etc.).

This article shows that such a conclusion is not always inevitable. Higher powered incentives *can* be restored by invoking selective audits and contingent monitoring. The idea is to make employees *wish* that their performance on softer tasks be audited, but to condition the occurrence of an audit on observing *high* output on those tasks that are easier to appraise. Under this scheme an employee would not increase effort on the tangible tasks, thereby raising the frequency of audits, without increasing as well his effort on the intangible tasks so that audit assessments are good. The structure of rewards creates a complementarity between efforts on the different tasks, which counteracts the substitution effect already present on the cost side. If the net effect is that efforts on all tasks are complements, then the usual conflict between tasks is lifted and the above argument in support of weaker incentives does not hold.

One legitimate question to be asked at this point is whether the present auditing scheme could effectively be used in real organizations. Its implementation seems indeed limited by the fact that audits are generally still costly: auditors need to be trained, their independence must be guaranteed, the scope and procedures of audits have to be defined and enforced, and completion of an audit takes time. First, let us emphasize that we are talking here about *internal audits*. The technology of such audits is now improving rapidly, thanks to recent advances in information technology, and in quality and environmental management [e.g., the ISO 9000 and ISO 14000 standards, and the EMAS (Environmental Management and Audit Scheme) standard in Europe]. Nowadays, internal audits are often routinely used to check compliance with policies and regulations, the safeguarding of assets, the economical and efficient use of resources, and established goals and objectives within the firm. They are increasingly considered an indispensable tool for control and enforcement of the policies established by senior management and corporate boards. It is likely that, in the near future, their cost will keep decreasing rapidly and that their scope will keep raising.<sup>14</sup> Second, let us point out that the term

---

14. See the actual assessments and studies by The Institute of Internal Auditors. According to its president (see William J. Bishop, 1997): "Technology has significantly changed the profession and practice of internal auditing as audit tools today are almost entirely automated and are rapidly growing more sophisticated."

“audit” in this article should not be taken too literally, but should be understood as referring to any kind of “contingent performance appraisal.” One possible concrete illustration of the present scheme could therefore be, for instance, the fast-track career paths frequently encountered in large firms. Fast-tracks allow employees to be considered earlier for promotion when they show outstanding performance on some of their preassigned tasks. Promotion is never a sure thing, however. Results on softer tasks, such as involvement in the internal working of the organization, assistance to teammates, and implementation of long-term goals, are first checked more closely to determine the candidate’s potential. In this context the article predicts that an employee who spends extra effort on regular duties in order to draw attention from his bosses would also work harder on those less tangible tasks that matter for promotion. To introduce complementarities between tasks, and thereby raise the power of incentives, might therefore be an implicit rationale for the use of fast tracks.

One drawback of the present scheme which may limit its application, however, is the very strong requirements it puts on the principal’s commitment ability. First, audits being costly, the principal should want to renege on auditing the agent’s performance on the remaining task even when she has to. Of course, the auditing scheme would not work if the principal could do so. One way to cope with this is to ensure that audits be triggered by signals that are objective, straightforward, and verifiable. If everybody (including the courts) acknowledges that the agent’s performance on the monitored task is outstanding, then the principal will keep her promise to audit the agent, in order to avoid losing face, damaging her reputation, and incurring direct penalties for breach of contract. Increasing the principal’s commitment to the above auditing scheme amounts to putting additional demands on the information technology associated with the monitored task. Second, audits being noisy, their conclusions are subject to manipulation and the parties involved might not accept them. One solution here is that principal and agent agree *ex ante* on verifiable standards for the practice of audits and the selection of auditors. This can be difficult to realize, but an important contribution of recognized professional and training institutions, and of international industry standards such as the ISO norms, is in fact to economize on reaching such an agreement.

One may finally wonder whether the present scheme can be extended to situations with more than two tasks. Suppose for instance that the agent’s job involves several tasks A, B, C, and so on. In this context an auditing scheme analogous to the one above would be the following. Let audits still be perceived *ex ante* by the agent as a reward, and let the agent be penalized when an audit reaches a bad conclusion. Now, let task A output be monitored; if it is high, then audit performance on task B; if the latter is also very good, then audit performance on task C; and so on. This scheme would also make the efforts on the various tasks complementary in the agent’s reward function. However, the range of circumstances where it would turn out to be optimal seems more limited.

## Appendix

*Statement.* The optimal contract with  $m_i \equiv 1$  is such that the difference  $u(w_{ij}) - u(w_{ij-1})$  is nonincreasing in  $i$  in the following cases:

- (a) with constant absolute risk aversion, that is, when  $u(z) = 1 - e^{-rz}$ ;
- (b) with constant relative risk aversion equal to 1, that is, when  $u(z) = \ln(1 + z)$ ;
- (c) when  $u(z) = (1 + z)^\alpha$  with  $0 < \alpha < 1/2$ .

*Proof.* First consider problem (4) with  $m_i \equiv 1$ , that is,

$$\begin{aligned} & \underset{w, a, b}{\text{maximize}} \sum_{i,j} p_i(a) q_j(b) (R(i, j) - w_{ij}), \text{ subject to:} \\ & \sum_{i,j} p'_i(a) q_j(b) u(w_{ij}) - c_a \geq 0 \\ & \sum_{i,j} p_i(a) q'_j(b) u(w_{ij}) - c_b \geq 0 \\ & \sum_{i,j} p_i(a) q_j(b) u(w_{ij}) - c(a, b) \geq U^*. \end{aligned} \tag{A1}$$

Let  $\gamma$ ,  $\lambda$ , and  $\mu$  be the Lagrange multipliers associated with the first, second, and third constraints, respectively. After a little bit of algebra, the first-order necessary (and sufficient) conditions for an optimal wage schedule are then given by

$$\forall i, j: \frac{1}{u'(w_{ij})} = \mu + \gamma \frac{p'_i(a)}{p_i(a)} + \lambda \frac{q'_j(b)}{q_j(b)}. \tag{A2}$$

Since the Lagrange multipliers are nonnegative and  $u$  is concave, the wage schedule  $w_{ij}$  must be nondecreasing in  $i$  and  $j$  by MLRP. Now, from Equation (A2), we have

$$\forall i, j: \frac{1}{u'(w_{ij})} - \frac{1}{u'(w_{ij-1})} = \lambda \left[ \frac{q'_j(b)}{q_j(b)} - \frac{q'_{j-1}(b)}{q_{j-1}(b)} \right]. \tag{A3}$$

Consider now the three cases of the statement.

- (a) If  $u(z) = 1 - e^{-rz}$ , note that  $1/u'$  is a convex increasing function. Therefore, since  $w_{ij}$  is nondecreasing, the differences  $w_{ij} - w_{ij-1}$ , and so  $u(w_{ij}) - u(w_{ij-1})$  as well, cannot increase in  $i$ .
- (b) When  $u(z) = \ln(1 + z)$ , we have that  $1/u'(z) = 1 + z$ . In this case, the differences  $w_{ij} - w_{ij-1}$  remain constant in  $i$ . Hence,  $u(w_{ij}) - u(w_{ij-1})$  must be decreasing with respect to  $i$  since  $u$  is concave.
- (c) Finally, if  $u(z) = (1 + z)^\alpha$  with  $0 < \alpha < 1/2$ , the function  $1/u'(z) = (1/\alpha)(1 + z)^{1-\alpha}$  is concave increasing. The differences  $w_{ij} - w_{ij-1}$  are then increasing in  $i$ , but it can be checked that they do not increase enough to make  $u(w_{ij}) - u(w_{ij-1})$  increase with  $i$  as well. ■

## References

- Anderson, Erin, and David Schmittlein. 1984. "Integration of the Salesforce: An Empirical Examination," 15 *RAND Journal of Economics* 385–95.
- Arrow, Kenneth J. 1965. *Aspects of the Theory of Risk Bearing*. Helsinki: Yrjö Jahnssoonin Säätiö.

- Baiman, Stanley, and Joel Demski. 1980. "Economically Optimal Performance Evaluation and Control Systems," 18 *Journal of Accounting Research* 184–220.
- Bishop, William G. 1997. "Technology: Changing the Rules for Business Controls and Internal Auditing," Nov/Dec *Business Credit* [also available on The Institute of Internal Auditors' website: www.theiia.org].
- Carroll, Christopher D., and Miles S. Kimball. 1996. "On the Concavity of the Consumption Function," 64 *Econometrica* 981–92.
- Debreu, Gérard, and Tjalling Koopmans. 1982. "Additively Decomposed Quasi-Convex Functions," 24 *Mathematical Programming* 1–38.
- DeGroot, Morris H. 1970. *Optimal Statistical Decisions*. New York: McGraw-Hill.
- Dixit, Avinash K. 1997. "Power of Incentives in Private versus Public Organizations," 87 *American Economic Review* 378–82.
- Dye, Ronald A. 1986. "Optimal Monitoring Policies in Agencies," 3 *RAND Journal of Economics* 339–50.
- Holmström, Bengt. 1979. "Moral Hazard and Observability," 10 *Bell Journal of Economics* 74–91.
- Holmström, Bengt, and Paul Milgrom. 1987. "Aggregation and Linearity in the Provision of Intertemporal Incentives," 55 *Econometrica* 303–28.
- . 1991. "Multitask Principal-Agent Analyses: Incentives Contracts, Asset Ownership, and Job Design," 7 *Journal of Law, Economics, & Organization* 24–52.
- Karp, Richard P. 1972. "Reducibility among Combinatorial Problems," in R. E. Miller and J. W. Thatcher, eds., *Complexity of Computer Computations*. New York: Plenum Press.
- Kimball, Miles S. 1990. "Precautionary Savings in the Small and in the Large," 58 *Econometrica* 53–73.
- Krueger, Alan B. 1991. "Ownership, Agency and Wages: An Examination of Franchising in the Fast Food Industry," 106 *Quarterly Journal of Economics* 75–101.
- Lazear, Edward P. 1995. *Personnel Economics*. Cambridge: MIT Press.
- MacDonald, Glen M., and Leslie M. Marx. 1998. "Adverse Specialization," Working Paper no. 98-05, Bradley Policy Research Center, University of Rochester.
- Miller, Gary J. 1992. *Managerial Dilemmas. The Political Economy of Hierarchy*. New York: Cambridge University Press.
- Mirrlees, James. 1976. "The Optimal Structure of Incentives and Authority within an Organization," 7 *Bell Journal of Economics* 105–31.
- Ross, Steven. 1973. "The Economic Theory of Agency: The Principal's Problem," 63 *American Economic Review* 134–39.
- Shavell, Steven. 1979. "Risk Sharing and Incentives in the Principal and Agent Relationship," 10 *Bell Journal of Economics* 55–73.
- Sinclair-Desgagné, Bernard. 1994. "The First-Order Approach to Multi-Signal Principal-Agent Problems," 62 *Econometrica* 459–65.
- Sinclair-Desgagné, Bernard, and H. Landis Gabel. 1997. "Environmental Auditing in Management Systems and Public Policy," 33 *Journal of Environmental Economics and Management* 331–46.
- Slade, Margaret E. 1996. "Multitask Agency and Contract Choice: An Empirical Exploration," 37 *International Economic Review* 465–86.
- Williamson, Oliver E. 1985. *The Economic Institutions of Capitalism*. New York: Free Press.