Australian School of Business

Working Paper

Never Stand Still   Australian School of Business

Australian School of Business Research Paper No. 2012 ECON 13

Moral Hazard with Discrete Soft Information

Guillaume Roger

Moral hazard with discrete soft information

Guillaume Roger *

The University of New South Wales

December 2011

Abstract

I study a simple model of moral hazard with soft information. The risk-averse agent takes an action and she *alone* observes the stochastic outcome; hence the principal faces a problem of *ex post* adverse selection. With limited instruments, the principal cannot solve these two problems independently. To accommodate *ex post* information revelation, he must distort the transfer schedule, as compared to the standard moral hazard problem. Then effort is implemented for a smaller set of parameters than in the standard problem. These results are robust and suggest high-power contracts may have to be revisited when information is soft.

Keywords: moral hazard, asymmetric information, soft information, contract, mechanism, audit. JEL Classification: D82.

---

*School of Economics, UNSW. Email: g.roger@unsw.edu.au. I specially thank Guido Friebel, Claudio Mezzetti, Mike Peters, Ralph Winter, my colleagues Bill Schworm, Chris Bidner, Carlos Pimienta and Suraj Prasad, and Luis Vasconcelos for helpful comments on this paper and on a broader project. I benefited from remarks in seminars at UNSW, U. of Sydney, U. of Queensland, U. of Adelaide, Goethe Universität, Universidade Nova de Lisboa, U. of Melbourne and conference participants at the Australasian Theory Workshop 2010, the IIOC 2010, the 9th Journées L.-A. Gerard-Varet and the 2010 EEA Congress . Financial support from the Australian School of Business at UNSW is gratefully acknowledged.

1 Introduction

In standard moral hazard problems the outcome of the agent's action is observable by the principal and may therefore (imperfectly) substitute itself for the non-observability of said action. Then a complete contract may be conditioned on the outcome. This is a convenient model and is useful to study the cost of moral hazard; but it does not necessarily fit many relevant situations. Performance may be difficult to observe, or its observation may be considerably delayed. Sometimes it is not observed at all: for example, an accounting report is *not* a direct observation of the state of an enterprise, but a *message*. In this paper, attention is paid to the case where neither the action, nor the outcome are observable by the principal. The information is therefore said to be *soft*, in that it is subject to manipulation on the part of the agent. Applications of this model are broad-ranging. For example, after hiring the CEO, a board often asks of him (her) to report his (her) results while on the job. A defense contractor may be asked to reveal its production cost after investing in an uncertain technology. In the optimal taxation problem, the agent may undertake some investment (education) to enhance her productivity, and then be asked to reveal the latter to the tax authority. Information manipulation exists in practice: Kedia and Philippon (2006) develop and test a model of earnings management (a euphemism for fraudulent accounting). They document how pervasive the practice is.

Bar for the role of soft information, the model mirrors that of a standard moral hazard framework. A risk-neutral principal delegates production to a risk-averse agent. The agent's action a governs the distribution of a stochastic outcome drawn from a discrete space, which she *alone* observes. In this construct, the principal is exposed to *ex ante* moral hazard and also faces a problem of adverse selection *ex post*. That information must therefore be elicited; that is, it must satisfy some *ex post* incentive constraints.[1] Because the principal otherwise observes nothing, the contract must include an audit and some (exogenously-given) penalty. The model is *not* reliant on endogenous penalties or rewards; that is, the principal possesses

[1]Throughout I will refer to "incentive constraints" as those addressing the adverse selection problem and "moral hazard" constraint as that dealing with the hidden action problem.

fewer instruments than in Kanodia (1985) or Mookherjee and Png (1989). This paucity of instruments induces a fundamental tension between *ex ante* effort provision, which requires a state-contingent compensation, and *ex post* information revelation, which is best addressed with a constant transfer.

This simple model delivers some important insights. First, absent some measure of *type separation*, the principal can only offer a trivial contract, in which the agent exerts no effort. Indeed, failing to separate is observationally equivalent to producing the same output. Therefore, in this model, at least *some* information must be revealed for the principal to be willing to implement the high action. Second, *any* information revelation requires the agent's compensation to be distorted as compared to the standard moral hazard model. This is necessary to simultaneously satisfy any of the *ex post* incentive constraints and the moral hazard constraint. The fundamental driving force behind these two observations is the conflict between *ex post* incentives for information revelation and *ex ante* effort incentives. The one implication of these two observations is that effort is more costly to the principal, and will therefore be implemented for a smaller set of parameters than in the standard problem. Incidentally, these distortions leave the agent's participation constraint slack (absent a participation fee) and thus generate an *ex ante* rent.[2]

Ignoring the trivial outcome of no effort, the subgame perfect equilibrium of this game is unique and induces truthful revelation. The compensation schedule is flatter than in the standard model. A steep transfer scheme is usually good to induce effort, but here it also generates incentives to misreport *ex post*. The distortions accommodate this new problem. The optimal contract is low(er)-powered, which suggests that high-power schemes such as options may be ill-suited when the information on which they are conditioned is soft. Real-life stories like the bankruptcy of Enron, for example, abound in support of this remark. A better audit technology mitigates the need for distortions but even if it is perfect, it never removes them. The reason is this. The standard second-best contract has a binding

[2]Because a participation fee would be sunk at the stage of information revelation the aforementioned distortions would remain.

participation constraint, for which the agent must be presented with negative *ex post* utility in the worse state(s). Here, even with a perfect *ex post* audit, the agent can always do better than accepting such a bad outcome by simply misreporting and facing a lottery over zero and some positive payoff. So adding this communication stage after the realisation of the state dramatically alters the nature of the contract.[3]

The work closest to this is Mookherjee and Png (1989), who combine a Grossman-Hart (1983) model with an *ex post* revelation mechanism. The agent's message conditions a payment to the principal and the probability of audit; that audit is perfect and fines or rewards (for truthful revelation) may be used. Only the latter are offered in equilibrium and they may be arbitrarily large, which may turn the principal into a source of money.[4] With these, truthful revelation obtains and has no bearing on the moral hazard problem because the principal has enough instruments to separate the *ex post* problem (information revelation) from the *ex ante* one (moral hazard). The goal of this paper is precisely to restore and study that connection. This model differs from their paper in two ways. First, there are no endogenous penalties for misreporting nor rewards for truthful revelation; the principal thus must do with fewer instruments. Truthful revelation obtain in my model because the principal distorts transfers. Second, the audit is imperfect. The technology is closer to one of sampling, which is what most real audits do, and has been modeled by Bushman and Kanodia (1996) or Demski and Dye (1999).

Gromb and Martimort (2007) use the same sequence of events as here, however they study the incentives of expert(s) to search and report information about *others* (an exogenous project), not themselves. To overcome the adverse selection problem, their incentive contract must be made state dependent although the expert(s) do(es) not exert any influence on it. A project's success is publicly observable, hence contractible. Levitt and Snyder (1997) develop a contracting model in which the agent receives an early (soft) signal about the likely success of the project, however the eventual outcome is fully observed by the principal, hence contractible. Here, information can only be observed, and reported, by the agent. To

[3]This differs from the problem of variance reduction of Holmström and Milgrom (1987).

[4]Mookherjee and Png's model yields a quirky byproduct: the agent strictly prefers being audited.

emphasize the point, the agent has no *ex ante* private information, which only emerges *ex post*. Green and Laffont (1986) study the principal-agent problem with "partially verifiable information" in the sense that the agent's message is constrained to lie in an arbitrary subset $M(\theta)$ of the type space Θ, which varies with the true state in a publicly known fashion. $M(\cdot)$-implementable mechanisms exist and need not elicit truth-telling. I discuss this some more in Section 5.

The balance of the paper is organised as follows. After introducing the model I present the problem of information revelation. Section 4 then analyses the optimal contract and Section 5 presents a discussion. I conclude in Section 6.

2 Model

A principal delegates a task to an agent. At cost $\psi(a) \geq 0$ the agent undertakes an action $a \in \{\underline{a}, \overline{a}\}$ where $\psi(\overline{a}) = \psi > \psi(\underline{a}) = 0$. This action yields a stochastic outcome $\theta \in \Theta \equiv \{\theta_1, \theta_2, \theta_3\}$. I make the substantive (and later discussed) assumption that $\theta_3 - \theta_2 = \theta_2 - \theta_1 = \Delta\theta > 0$. Let $\overline{\pi}_i \equiv \Pr(\theta_i | a = \overline{a})$ and $\underline{\pi}_i \equiv \Pr(\theta_i | a = \underline{a})$ denote the probabilities of each outcome conditional on the agent's action, with $\overline{\pi}_i > \underline{\pi}_i$; I also suppose that $\pi_i(\cdot|a)$ satisfies the MLRP (i.e. $\Delta\pi_i/\overline{\pi}_i$ increases in θ). The agent's net utility is given by $u(t, a) \equiv v(t_i) - \psi(a)$, where $v(\cdot)$ is an increasing, concave function with $v(0) = 0$. The agent *alone* observes the outcome θ and reports a message $m \in \Theta$ to the principal, whereupon she receives the transfer $t_i(m)$. The principal can commit to the contract and his net payoff reads $S(t; \theta) = \theta_i - t_i$. If the true state θ were observable by the principal, this construct would be a moot point and would collapse to the textbook moral hazard problem.

Inducing effort requires $t_3 \geq t_2 \geq t_1$ with at least one strict inequality. Therefore it is immediate that absent any other instruments, the agent pools her messages to θ_3 regardless of the state (the principal entirely lacks *ex post* observability). A necessary element of any non-trivial contract is to restore at least some *ex post* observability, which I do by introducing an *ex post* audit. It is exogenously given in this model, costless (and therefore always run), but imperfect.[5] With some probability $p(m-\theta)$, the agent's deception is uncovered and she receives zero. Any penalty that is not too large may work; so conditional on accepting the premise that the penalty must be bounded, choosing zero is immaterial (see the Discusion).[6] The function $p(\cdot)$ maps into $[0,1]$; it is taken to be symmetric and such that $p(0)=0$.[7] It is easy to show that $p(\cdot)$ must be increasing to be useful, which I therefore impose. I also let $p(2x) \geq 2p(x)$. With this, the agent has *ex post* expected utility

$$U \equiv (1-p(m-\theta))v(t(m)),$$

which she seeks to maximise by choice of the message $m \in \Theta$. The timing is almost standard:

1. The principal offers a contract $\mathcal{C} = \langle t(m), \Theta, p(\cdot) \rangle$ consisting of a transfer, a message space and an (exogenous) audit probability;
2. The agent accepts or rejects the contract. If accepting, she also chooses an action a;
3. Action a generates an outcome $\theta \in \Theta$ observed only by the agent;
4. The agent report a message $m \in \Theta$;
5. Audit occurs;
6. Transfers are implemented and payoffs are realised.

The solution of this problem in a subgame perfect equilibrium of the game just described.

[5]In a separate paper I allow the audit function $p(\cdot)$ to be endogenous. Then it interacts with other variables of the contract.

[6]A very large penalty would yield truthful revelation at no cost, as in Mookherjee and Png (1989).

[7]This capture the idea that the audit is a sampling process, as real financial audits are.

3 Information revelation

After she has taken some action a (now sunk), the agent maximises U by choice of a message m. A mechanism is truthful if and only is the following constraints are satisfied:-

$$v(t_1) \geq (1 - p(\Delta\theta))v(t_2) \tag{3.1}$$

$$v(t_1) \geq (1 - p(2\Delta\theta))v(t_3) \tag{3.2}$$

$$v(t_2) \geq (1 - p(\Delta\theta))v(t_3) \tag{3.3}$$

$$v(t_2) \geq (1 - p(\Delta\theta))v(t_1) \tag{3.4}$$

$$v(t_3) \geq (1 - p(2\Delta\theta))v(t_1) \tag{3.5}$$

$$v(t_3) \geq (1 - p(\Delta\theta))v(t_2) \tag{3.6}$$

These constraints do not yield the standard implementability condition, as can be verified by adding them up pairwise. For example, add (3.1) and (3.4) to find $1 \geq (1 - p(\Delta\theta))$, which is trivially true and uninformative as to the shape of the transfer function. The system (3.1)-(3.6) forms the basis of the first claim. Because $t_3 \geq t_2 \geq t_1$ and $p(\cdot) \geq 0$, only (3.1)-(3.3) are relevant.

Lemma 1 *There exist transfers $t_3 \geq t_2 \geq t_1$ such that constraints (3.1)-(3.6) hold. Whenever the local constraints (3.1) and (3.3) are satisfied, the global constraint (3.2) is necessarily slack. Whenever the global constraint (3.2) binds at least one of the local constraints fails.*

This existence result remains silent as to optimality and does not imply that transfers satisfying (3.1)-(3.6) solve the principal's problem. Truthful revelation needs not be optimal, in particular because it necessarily generates an *ex ante* rent for the agent. Indeed, $t_3 > 0$ is required to induce participation with any effort but by (3.2), $t_1 > 0$ as well (and so $t_2 > 0$ too). However from Holmström (1979) we know that some t_i must be negative for the participation constraint to bind. Here the principal's choice of offering a contract such that truthful revelation obtains induces a lower bound on the transfers, akin to a limited liability constraint–hence the rent.

4 The optimal contract

There always exists a trivial contract, in which the low action is sought from the agent. When $a = \underline{a}$, the principal needs only set $t_i = 0 \; \forall i$ (which elicits truth-telling *ex post*). I am interested in equilibria where effort is implemented. Two cases of interest arise. In the first one, truthful revelation is elicited *ex post*, which may come at a cost to the principal. In the second case, the principal may not seek to satisfy the *ex post* incentive constraint because it is too costly. Let $\varphi \equiv v^{-1}(\cdot)$ denote the inverse function of the agent's utility, and $v_i = v(t_i)$ for some t_i. The goal of the next two subsections is to compute the cost of either contract, which is the outcome of the Nash equilibrium of each subgame.

4.1 Truth-telling equilibrium

A truth-telling equilibrium is one where all the *ex post* incentive constraints (3.1)-(3.3) are satisfied. The principal seeks to solve

Problem 1

$$\max_{v_i \geq 0} \sum_i \overline{\pi}_i \left[\theta_i - \varphi(v_i)\right]$$

s.t. (3.1)-(3.3) and

$$\sum_i \Delta\pi_i v_i \geq \psi \tag{4.1}$$

$$\sum_i \overline{\pi}_i v_i \geq \psi \tag{4.2}$$

The last two inequalities are the usual moral hazard and participation constraints. Following Lemma 1, only (3.1) and (3.3) may bind, so attach multipliers γ_1, γ_2 to these constraints and λ and μ to each of (4.1) and (4.2), respectively. Lemma 1 also tells us that $\mu = 0$ for constraints (3.1)-(3.3) to be satisfied. The next two lemmata inform us more precisely as to how these constraints conflict with the moral hazard problem.

Lemma 2 *Suppose $\mu, \lambda > 0$ (as in the standard moral hazard problem), then at least two of (3.1)-(3.3) must be violated.*

Hence there cannot be an equilibrium in which the standard solution of the moral hazard problem also accommodates the *ex post* information revelation problem. Further, for the constraints (3.1)-(3.3) to hold, either (4.1) fails or (4.2) must be slack. More precisely,

Lemma 3 *Suppose $\mu > 0$ and (3.1), (3.3) are satisfied, then the moral hazard constraint (4.1) is violated.*

That is, either there is no truthful revelation *ex post* (by Lemma 2), or no effort can be induced without affording the agent a rent (by Lemma 3). Therefore I can restrict attention to the set of utilities v_i such that (4.2) is slack. Taking (3.1) and (3.3) binding, the transfers must satisfy $v_1 = (1-p(\Delta\theta))v_2$ and $v_2 = (1-p(\Delta\theta))v_3$. Define further $\overline{\Pi} = \overline{\pi}_3 + (1-p)\overline{\pi}_2 + (1-p)^2\overline{\pi}_1$ and $\underline{\Pi} = \underline{\pi}_3 + (1-p)\underline{\pi}_2 + (1-p)^2\underline{\pi}_1$. With this in hand,

Proposition 1 *The lowest-cost truth-telling equilibrium in which the agent is induced to exert effort entails*

$$v_3^T = \frac{\psi}{\overline{\Pi} - \underline{\Pi}}$$

determined by a binding moral hazard constraint (4.1), and $v_1^T, v_2^T > 0$ determined by (3.1) and (3.3), both binding.

More can be said. The necessary and sufficient first-order conditions of Problem 1 are:-

$$\varphi'(v_1) - \frac{\gamma_1}{\overline{\pi}_1} = \mu + \lambda\frac{\Delta\pi_1}{\overline{\pi}_1} \tag{4.3}$$

$$\varphi'(v_2) - \frac{\gamma_2 - \gamma_1(1-p)}{\overline{\pi}_2} = \mu + \lambda\frac{\Delta\pi_2}{\overline{\pi}_2} \tag{4.4}$$

$$\varphi'(v_3) + \frac{\gamma_2(1-p)}{\overline{\pi}_3} = \mu + \lambda\frac{\Delta\pi_3}{\overline{\pi}_3} \tag{4.5}$$

The MLRP ensures these conditions are not vacuous; the next claim follows immediately.

Proposition 2 *The schedule is flatter (than under the standard moral hazard problem). v_1^T solving (4.3) exceeds the standard level, and v_3^T solving (4.5) is lower than the standard level. v_2 is ambiguous.*

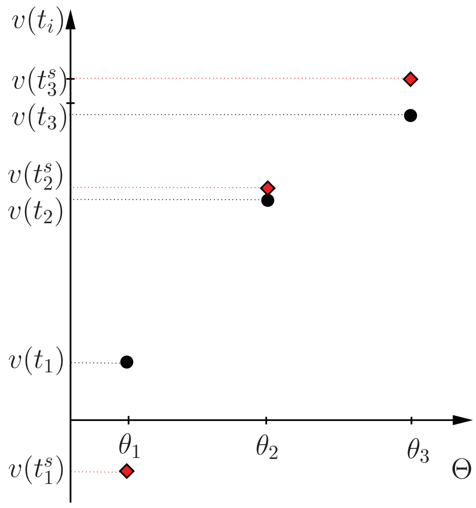

Figure 1: Optimal transfers are distorted (black dots).

This result owes to the fundamental tension between *ex post* incentive compatibility, best satisfied with constant transfers, and *ex ante* effort incentives, best addressed with a compensation conditioned on performance. The distortions tilt the schedule and are such that (3.1) and (3.3) are just binding. The contract is low(er)-powered to accommodate *ex post* information manipulation. This is shown in Figure 1. Finally,

Corollary 1 *The principal induces costly effort if and only if* $\sum_i \Delta\pi_i\theta_i \geq \psi\frac{\overline{\Pi}}{\overline{\Pi}-\underline{\Pi}} > \psi$ *i.e. the agent receives an* ex ante *rent* $U^T = \psi\frac{\underline{\Pi}}{\overline{\Pi}-\underline{\Pi}} > 0$.

the proof of which is obvious and therefore omitted. Consequently the high action is implemented for a narrower set of parameters than under the standard moral hazard problem, which is costly to the principal. The rent is is excess of the standard risk-premium the principal must pay (to partially insure the agent).

4.2 No truthful revelation

Because combining *ex ante* effort incentives and *ex post* truthful revelation is costly, the principal may choose an alternative: truthful revelation may purposefully not be sought,

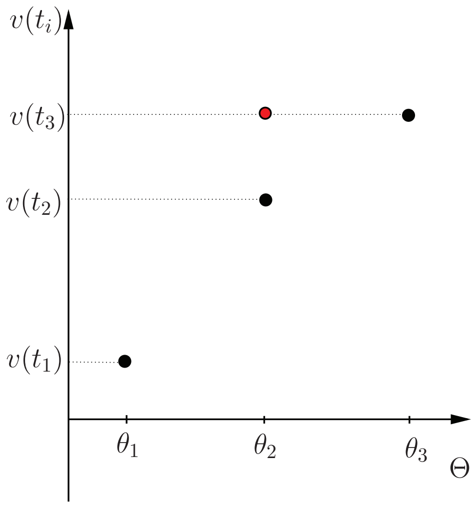

Figure 2: Arbitrary example of (3.3) failing, θ_2 pools with θ_3 (red dot).

which may make him better off. In this equilibrium, transfers are such that at least one of the *ex post* incentive constraints is violated, as in Figure 2.

Even when the agent sends a message that is not truthful, the principal does not update his beliefs as to the true state of the world because he has committed to the contract.[8] I argue next that it then follows that at least some of the incentive constraints (3.1) to (3.3) must hold. If *all* constraints were to fail, the agent could only report θ_3, whereupon the principal would pay t_3. But then there would be no incentive for the agent to exert any effort, so the principal would offer only $\underline{t}_i = 0$. Therefore, some measure of type *separation* is a *necessary* condition for the principal to want to induce action $\overline{a}$. One must also note that since some incentive constraint will fail (by design), (3.2) can no longer be ignored. However it is still true that there cannot be an equilibrium in which only (3.2) is violated, because (3.1) and (3.3) imply (3.2). Conversely, if (3.1) and (3.3) fail, it does not imply that (3.2) does. The principal's program is

[8]The principal has no further move in the game, so updating is a moot point. In particular, there is no renegotiation.

Problem 2

$$\max_{v_i \geq 0} \sum_i \overline{\pi}_i \theta_i - \sum_i \rho_i \varphi(v_i)$$

s.t. (3.1)-(3.3) and

$$\sum_i \Delta\rho_i v_i \geq \psi \tag{4.6}$$

$$\sum_i \overline{\rho}_i v_i \geq \psi \tag{4.7}$$

where ρ_i denotes the probability of receiving a report θ_i when the agent pools states (since some incentive constraint fails). The exact definition of ρ_i depends on the choice of pooling, that is, on which of the incentive constraints fail(s). At face value there are many combinations to consider; fortunately the next two lemmata reduce the set of cases to investigate.

Lemma 4 *The principal does not offer a contract in which the agent exerts effort such that (3.2) and (3.3) are violated.*

In this case the agent pools her message at θ_3 and no separation obtains. It is also true that

Lemma 5 *The principal does not offer a contract in which the agent exerts effort and any of*

1. *(3.1) and (3.2) or;*
2. *(3.1) and (3.3) or;*
3. *only (3.3);*

are violated.

Thus the only viable case when the principal's contract is such that it does not induce the agent to truthfully reveal her information *ex post* requires (3.1) to be violated. And here too the participation constraint must be left slack in order to satisfy the moral hazard constraint.

Lemma 6 *There is no equilibrium such that the moral hazard constraint (4.6) is satisfied, the participation constraint (4.7) binds, and only (3.1) is violated*

Therefore Lemma 1 and Lemma 6 together imply that in *any* equilibrium the agent receives an *ex ante* rent. With only one case to study and a slack constraint (4.7), Problem 2 becomes

Problem 3

$$\max_{v_i \geq 0} \sum_i \overline{\pi}_i \theta_i - [\overline{\pi}_3 \varphi(v_3) + (1-\overline{\pi}_3)\varphi(v_2)]$$

s.t. (3.2),(3.3) and

$$\Delta\pi_3(v_3 - v_2) \geq \psi \tag{4.8}$$

$$\overline{\pi}_3 v_3 + (1-\overline{\pi}_3)v_2 > \psi \tag{4.9}$$

where (4.9) is the relevant form of the participation constraint (4.7). Attach multipliers γ_2, γ_3 to (3.3), (3.2), and the usual λ to (4.8).

Proposition 3 *The least-cost non-truthful equilibrium in which the agent is induced to exert effort entails*

$$v_3^{NT} = \frac{\psi}{\Delta\pi_3 p(\Delta\theta)}; \quad v_2^{NT} = (1-p(\Delta\theta))v_3^{NT}; \quad v_2^{NT} = 0$$

determined by a binding moral hazard constraint (4.8) and a binding incentive constraint (3.3), at cost $C^{NT} = \frac{\psi}{\Delta\pi_3 p(\Delta\theta)}[1 - p(\Delta\theta)(1-\overline{\pi}_3)]$

Here v_1 is set to 0 without loss of generality; anything negative fails (3.1) and (3.2) as well but is irrelevant thanks to the agent's deviation option to misreport, while a positive utility level may (at least weakly) dilute the effort incentives. Because of the distortions the high action is implemented for a smaller set of parameters than in the standard case. To complement Proposition 3 observe that the agent receives an *ex ante* rent $U^{NT} = \frac{\psi}{\Delta\pi_3 p}[1 - p(\Delta\theta)(1-\underline{\pi}_3)] > 0$.

The first-order conditions of Problem 3 read:-

$$\begin{aligned} (1-\overline{\pi}_3)\varphi'(v_2) - \gamma_2 &= -\lambda\Delta\pi_3 \qquad (4.10) \\ \overline{\pi}_3\varphi'(v_3) + [(1-p(\Delta\theta)\gamma_2 + (1-p(2\Delta\theta))\gamma_3] &= \lambda\Delta\pi_3 \qquad (4.11) \end{aligned}$$

and directly lead to

Corollary 2 *The compensation schedule is flatter then in the standard moral hazard problem.*

4.3 Subgame perfect equilibrium

Collecting the results of the analysis of the two preceding sections it immediately follows that:

Proposition 4 *If he wants to induce the high action, the principal prefers to offer a truth-telling contract, i.e.* $\frac{\overline{\Pi}}{\overline{\Pi}-\underline{\Pi}} < \frac{1-p(\Delta\theta)(1-\overline{\pi}_3)}{\Delta\pi_3 p(\Delta\theta)}$.

This result is quite intuitive. The only non-truthful equilibrium is one where Constraint (3.1) fails and the θ_1 agent pools at θ_2. This increases the expected cost of the high action compared to the truthful equilibrium, without any effect on the effort level (and therefore on the distribution $\overline{\pi}_i$). This latter features owes to the binary nature of the action and so may be an artefact of this simple model. I offer a more comprehensive discussion in Section 5.

Altering any of the technology ψ or information structure $\pi(.|a)$ produces the same effects as in the standard moral hazard model. Of more interest are comparative statics with respect to the audit technology.

Corollary 3 *1.* $\frac{\partial U^T}{\partial p} < 0$

2. $\frac{\partial v_3^T}{\partial p} > 0$

Improving the audit technology tilts back the compensation schedule toward a steeper slope: the agent's utility in the good state increases. But it also eases the incentive constraints (3.1)-(3.3). The net effect is a decrease the agent's expected rent. There is a limit to this however. In this model even a perfect audit technology does not enable the principal to implement the standard moral hazard schedule. Consider (3.1) and (3.3) and suppose $p(\Delta\theta) = 1$. Then truthful revelation still requires $v_1 \geq 0$ and $v_2 \geq 0$, and since $v_3 > 0$, the participation constraint still fails to bind.[9] The reason for this seemingly counterintuitive result is that the zero-penalty acts like an implicit limited liability constraint. That is, no utility level $v_i < 0$ will ever be implemented in this model because the agent can always report anything else and

[9]In the standard model one must have at least $v_1 < 0$ to have a binding participation constraint.

face the lottery $\{p, 1 - p(\Delta\theta)\}$ over 0 and v_{i+1}, which has a strictly positive expected value. Thus the principal's reliance on a message from the agent alters the game dramatically.

Last, it is immediate that *sign* $dp/d\Delta\theta =$ *sign* $d\Delta\theta$ so that Corollary 3 extends to the type space. Increasing the distance between types (and therefore between messages) renders misreporting more hazardous here.

5 Discussion

5.1 Other penalties

This paper purposefully departs from optimal penalties because they allow for the separation of the *ex ante* and *ex post* problems (as in Mookherjee and Png (1989)). Suppose however that some other bounded penalty $-l < 0$ could be imposed on the agent, then the constraints (3.1)-(3.3) would be modified as follows: $v_i \geq [1 - p(\Delta\theta)]v_{i+1} - pl$ and $v_1 \geq [1 - p(2\Delta\theta)]v_3 - pl$. These are easier to satisfy than the current constraints but as long as l is not too large, the problem remains in essence the same.

It can also be verified that the Maximum Punishment Principle (Baron and Besanko (1984)) holds in this model because the audit does not return false negatives. Therefore nothing is gained (there are strict losses) from conditioning the penalty on the offence.

5.2 Participation fees

To avoid leaving any rent to the agent the principal could consider imposing an *ex ante* participation fee ϕ, so that $\mu > 0$. Then a contract includes a tariff (ϕ, t). But because ϕ is sunk, the incentive constraints (3.1)-(3.6) remain essentially the same (up to the levels of v_i). So although the rent may be extracted from the agent, the transfers t_i still have to be distorted to satisfy information revelation. Thus effort is still induced for a smaller set of parameters and welfare losses ensue.

5.3 Relation to M-implementability (Green and Laffont (1986))

These authors study the implementability of a social choice function when the agent may report a message from a set $M(\theta) \subset \Theta$, where $M(\cdot)$ is *exogenous* and publicly known. There is scope for misreporting in that the mapping $m(\cdot)$ is a correspondence. This clearly does not apply in the truthful equilibrium of this model, but it may in the alternative (which remains a Nash equilibrium of the game). Green and Laffont (1986) provide a necessary and sufficient condition – called the nested range condition (NRC) – for the agent to report her information truthfully.

The NRC does not hold in this model, although it corresponds to a game of of "unidirectional distortions with an ordered space" (to use their words) – example a(2) in Green and Laffont. Indeed, the NRC requires $M(\theta_3) = \{\theta_3\}$, $M(\theta_2) = \{\theta_2, \theta_3\}$, $M(\theta_1) = \{\theta_1, \theta_2, \theta_3\}$. In contrast, Constraints (3.2), (3.3) holding and (3.1) failing imply $M(\theta_3) = \{\theta_3\}$, $M(\theta_2) = \{\theta_2\}$, $M(\theta_1) = \{\theta_2\}$, whence $\theta_3 \notin M(\theta_1)$. This violates the definition of the NRC. Notice further that the sets $M(\theta_i)$ in this paper are endogenous, unlike in Green and Laffont (1986).

5.4 Separation versus truth-telling

That truthful revelation obtains in equilibrium likely owes to the combination of a binary action space with a discrete type space, which renders $p(m-\theta)$ "large" for any deviation (so misreporting has a high expected cost). In fact truthful revelation is a nice by-product but it is not essential. Indeed, in a separate paper with continuous actions and types, I show that truth-telling can never be elicited. This is because continuous spaces allow for arbitrarily small misreporting, and a small deviation may be optimal for some types. What is important for the provision of effort incentive is not truthful revelation, but *type separation*. In the extreme, if all types face incentives such that they all report the same message the agent has no incentive to expend any effort. That is, pooling stifles effort. So, that is not misreporting *per se* that deters *ex ante* incentives, but pooling. As we saw in Section 4, providing the agent with effort incentives may come at different costs when she is truthful and not.

5.5 Monitoring

Should the principal audit the agent's report of an *outcome* (θ), or should he somehow gather information about the action (a) – that is, monitor the agent? In the latter case, the agent is paid according to her action, not the outcome. Then the information revelation problem is moot and the risk-neutral principal bears all the risk. Implicitly in this paper it is presumed that monitoring is either too costly or outright impossible.

6 Conclusion

When the principal to a contract fraught with moral hazard also fails to observe any of the outcomes, he faces adverse selection *ex post*. With limited instruments, eliciting this private information requires a distortion of the compensation structure; it induces a flatter transfer scheme. This is a low(er)-power contract than in the standard moral hazard problem. This distortion has a bearing on the *ex ante* effort incentives. It is socially costly in that the high action can be implemented for a strictly smaller set of parameters than in the standard case. These results obtain because of the fundamental tension between effort provision and information revelation, which require different instruments. In this model, inducing *any* information revelation *ex post* generates an *ex ante* rent to the agent (but this is not an essential feature). For practitioners these results suggest that high-power contracts may not be adequate when the information they depend on can be manipulated.

Truthful revelation obtains in the unique (subgame perfect) equilibrium of this game, as it does in the Mookherjee and Png (1989), however for very different reason and with different consequences. In particular, *ex post* information revelation has no bearing on the *ex ante* moral hazard problem in Mookherjee and Png (1989). In contrast, information revelation can only be obtained with a distortion of the transfers, and therefore of the effort decision.

In a more general model, truthful revelation may not obtain. This more general model should allow for at least two important modifications: (i) richer and more flexible messages for the agent (i.e. relaxing $\Delta\theta = \theta_2 - \theta_1 = \theta_3 - \theta_2$) and (ii) afford the principal to choose his audit technology. However I conjecture that as long as instruments remain limited, the message of this paper essentially remains: the option to misreport *ex post* induces *ex ante* distortions.

7 Appendix: Proofs

Proof of Lemma 1: Suppose (3.1) and (3.3) are satisfied (either strictly or with some slack), then $v(t_1) \geq (1 - p(\Delta\theta))^2 v(t_3)$. Since $(1 - p(\Delta\theta))^2 > 1 - p(2\Delta\theta)$, Condition (3.2) is necessarily slack. To show existence, take (3.1) and (3.3) binding. Then we have $v(t_3) > (1 - p)v(t_3) = v(t_2) > (1 - p)^2 v(t_3) = v_1$. For the last statement, take (3.2) binding. Then $v(t_1) = (1 - p(2\Delta\theta))v(t_3) \geq (1 - p(\Delta\theta))v(t_2)$ by (3.1) (or $\geq (1 - p(\Delta\theta))v(t_1)$ by (3.3)). Either way it follows that $1 - p(2\Delta\theta))v(t_3) \geq (1 - p(\Delta\theta))^2 v(t_3)$ for both (3.1) and (3.3) to hold. This contradicts $(1 - p(\Delta\theta))^2 > 1 - p(2\Delta\theta)$. ■

Proof of Lemma 2: $\mu, \lambda > 0 \Leftrightarrow \sum_i \overline{\pi}_i v_i = \psi = \sum_i \Delta\pi_i v_i \Leftrightarrow \sum_i \underline{\pi}_i v_i = 0$. Since $v_3 \geq v_2 \geq v_1$ by MLRP and $\sum_i \overline{\pi}_i v_i = \psi > 0$, we must have $v_3 > 0 > v_1$. Therefore (3.2) cannot hold. Further, since $v_2 \geq v_1$, (3.1) must also be violated regardless of whether $v_2 \geq 0$ or $v_2 < 0$. Clearly in the later case (3.3) also fails to hold. ■

Proof of Lemma 3: $\mu > 0 \Leftrightarrow \sum_i \overline{\pi}_i v_i = \psi$ and when (4.1) holds, $\sum_i \overline{\pi}_i v_i \geq \psi + \sum_i \underline{\pi}_i v_i$. But but by (3.1)-(3.3) and MLRP, $\sum_i \underline{\pi}_i v_i > 0$. Then (4.1) implies

$$\begin{aligned}
\sum_i \overline{\pi}_i v_i &\geq \psi + \sum_i \underline{\pi}_i v_i \\
\psi &\geq \psi + \sum_i \underline{\pi}_i v_i \\
0 &\geq 0 + \sum_i \underline{\pi}_i v_i > 0
\end{aligned}$$

which is an obvious contradiction. So (4.1) must be violated. ■

Proof of Proposition 1: Set $\mu = 0$ and sum (4.3)-(4.5) to find

$$-\mathbb{E}_\Theta[\varphi'(v_i)] + (\gamma_1 + \gamma_2)p = 0$$

so that at least one of γ_1, γ_2 is strictly positive. Take the moral hazard constraint (4.1) and the *ex post* incentive constraints (3.1) and (3.3) binding to obtain v_3 (as in the text). Compute the cost of inducing effort as $C = \psi\frac{\overline{\Pi}}{\overline{\Pi}-\underline{\Pi}}$ and the rent by subtracting the cost of effort ψ. To show this must be the lowest-cost contract, observe that v_3 must also solve

$$\varphi'(v_3) = \lambda\frac{\Delta\pi_3}{\overline{\pi}_3} - \frac{\gamma_2(1-p)}{\overline{\pi}_3} \tag{7.1}$$

Since $\varphi'(\cdot)$ is increasing, if $\gamma_2 = 0$ then $v_3 > \frac{\psi}{\overline{\Pi}-\underline{\Pi}}$ and $v_2 > (1-p)v_3 > (1-p)\frac{\psi}{\overline{\Pi}-\underline{\Pi}}$. It then follows that $v_1 \geq (1-p)v_2 > (1-p)^2 v_3 > (1-p)^2\frac{\psi}{\overline{\Pi}-\underline{\Pi}}$. A similar argument can be constructed for the case $\gamma_1 = 0$ by using (4.4) instead. ■

Proof of Proposition 2: In the standard problem (4.3)-(4.5) read

$$\begin{aligned}
\varphi'(v_1) &= \mu + \lambda\frac{\Delta\pi_1}{\overline{\pi}_1} && (7.2)\\
\varphi'(v_2) &= \mu + \lambda\frac{\Delta\pi_2}{\overline{\pi}_2} && (7.3)\\
\varphi'(v_3) &= \mu + \lambda\frac{\Delta\pi_3}{\overline{\pi}_3} && (7.4)
\end{aligned}$$

Increase v_i $\forall i$ by some arbitrarily small amount $\varepsilon > 0$, so that $\mu = 0$ as well but the cost of the contract comes within ε of the optimum. Then compare each of (7.2)-(7.4) to (4.3)-(4.5), recalling that $\varphi(\cdot)$ is increasing convex. ■

Proof of Lemma 4: Suppose these constraints fail for some contract, then the agent only ever reports θ_3 and receives $\varphi(v_3)$ with probability 1. But then there is no incentive for effort. ■

Proof of Lemma 5:

1. The first case is not so obvious because it allows for $v_1 < 0 < v_2 < v_3$. But then the agent pools at θ_3 when observing θ_1. So Problem 2 becomes

   Problem 4

   $$\max_{v_i \geq 0} \sum_i \overline{\pi}_i\theta_i - [\overline{\pi}_2\varphi(v_2) + (1-\overline{\pi}_2)\varphi(v_3)]$$

   s.t. (3.3) and

   $$\Delta\pi_2(v_2 - v_3) \geq \psi \tag{7.5}$$

$$\overline{\pi}_2 v_2 + (1-\overline{\pi}_2) v_3 \geq \psi \tag{7.6}$$

Any effort on the part of the agent requires $v_3 \geq v_2$, which immediately violates the moral hazard constraint. The principal will never offer such a contract.

2. In the second case, v_1 again is "off equilibrium" because the agent reports θ_2 if observing θ_1 and otherwise pools at θ_3. The moral hazard constraint (7.5) rewrites $\Delta\pi_1(v_2 - v_3) \geq \psi > 0$, which is a contradiction again. The principal will never offer such a contract.

3. In the last instance, v_2 is also "off equilibrium" in the sense that the agent will always pool at θ_3 instead of reporting θ_2. Then the moral hazard constraint (7.5) becomes $\Delta\pi_1(v_1 - v_3) \geq \psi > 0$, which can never be satisfied. The principal will never offer such a contract either.

■

Proof of Lemma 6: Suppose (3.1) fails but (3.2) and (3.3) hold. θ_1 is never reported to the principal. By (4.7) and MLRP, $v_3 > v_2 > 0$. Regardless of the exact definition of ρ_i, take (4.7) binding then (4.6) is necessarily violated. ■

Proof of Proposition 3: Adding the first-order conditions, one finds

$$-\mathbb{E}_\Theta[\varphi'(v_i)] + p(\Delta\theta)\gamma_2 - (1 - p(2\Delta\theta))\gamma_3 = 0$$

whence $\gamma_2 > 0$ necessarily, and $v_2^{NT} = (1-p)v_3^{NT}$. γ_3 can be anything: since θ_1 is never reported, t_1 is never paid so any transfer satisfying (3.2) but not (3.1) will do. Combining the binding moral hazard constraint with $v_2^{NT} = (1-p)v_3^{NT}$ gives v_3^{NT}. Next compute the principal cost of inducing effort

$$\begin{aligned} C^{NT} &= \overline{\pi}_3 v_3^{NT} + (1-\overline{\pi}_3)v_2 \\ &= \overline{\pi}_3 \frac{\psi}{\Delta\pi_3 p(\Delta\theta)} + (1-\overline{\pi}_3)(1-p)\frac{\psi}{\Delta\pi_3 p(\Delta\theta)} \\ &= \frac{\psi}{\Delta\pi_3 p(\Delta\theta)}[1 - p(\Delta\theta)(1-\overline{\pi}_3)] \end{aligned}$$

and subtract the agent's effort cost ψ to find the rent U^{NT}. ■

Proof of Proposition 4: The principal chooses to offer a truth-telling contract if and only if $\psi\frac{\overline{\Pi}}{\overline{\Pi}-\underline{\Pi}} \leq \psi\frac{1-p(\Delta\theta)(1-\overline{\pi}_3)}{\Delta\pi_3 p(\Delta\theta)}$ (the cost of inducing the high-action is weakly lower). This condition rewrites

$$\frac{\overline{\pi}_3+(1-p)\overline{\pi}_2+(1-p)^2\overline{\pi}_1}{\Delta\pi_3+(1-p)\Delta\pi_2+(1-p)^2\Delta\pi_1} = \frac{1-p[\overline{\pi}_2+(2-p)\overline{\pi}_1]}{\Delta\pi_3+(1-p)\Delta\pi_2+(1-p)^2\Delta\pi_1} \leq \frac{1-p[\overline{\pi}_1+\overline{\pi}_2]}{\Delta\pi_3 p}$$

and is obviously always verified. ■

References

[1] Baron D. and David Besanko"Regulation, Asymmetric Information, and Auditing." *The RAND Journal of Economics*, Vol. 15, No. 4, pp. 447-470

[2] Bushman, R. and Chandra Kanodia (1996) "A Note on Strategic Sampling in Agencies." *Management Science*, Vol.42, pp. 151-156

[3] Demski, J. and Ronald Dye (1999) "Risk, Returns and Moral Hazard" *Journal of Accounting Research*, Vol. 37, No. 1, pp. 27-55

[4] Green, J. and Jean-Jacques Laffont (1986) "Partially verifiable information and mechanism design." *The Review of Economic Studies*, Vol. LIII, pp. 447-456

[5] D. Gromb and David Martimort (2007), "Collusion and the organization of delegated expertise" *Journal of Economic Theory*, Vol. 137 (1), pp. 271-299

[6] Grossman, S. and Oliver Hart (1983), "An Analysis of the Principal-Agent Problem" *Econometrica*, vol. 51(1), pages 7-45.

[7] Holmström, B. (1979) "Moral hazard and observability." *The Bell Journal of Economics*, Vol. 10, pp. 74-91

[8] Holmström, B. and Paul Milgrom (1987) "Aggregation and Linearity in the Provision of Intertemporal incentives." *Econometrica*, Vol. 55, pp. 303-328

[9] Kanodia, C. (1985) “Stochastic Monitoring and Moral Hazard.” *Journal of Accounting Research*, Vol.23, pp. 175-194

[10] Kedia, S. and Thomas Philippon (2006) “The economics of fraudulent accounting”, *NYU working paper*

[11] Levitt, S. and Christopher Snyder (1997) “Is no News Bad News? Information Transmission and the Role of ”Early Warning” in the Principal-Agent Model.” *Rand Journal of Economics*, Vol. 28 (4), pp. 641-661

[12] Mookherjee D. and Ivan Png (1989) “Optimal Auditing, Insurance, and Redistribution.” *The Quarterly Journal of Economics*, vol. 104(2), pp. 399-415