Moral hazard with soft information

Guillaume Roger *

The University of New South Wales

PRELIMINARY AND INCOMPLETE

August 2010

Abstract

We study a model of moral hazard with soft information: the agent takes an action and she alone observes the stochastic outcome, hence the principal faces a problem of ex post adverse selection. High-power contracts may not be appropriate when information is soft. Presuming of the Revelation Principle, the optimal (direct) mechanism with audit requires a two-part tariff to be offered to the agent. The fixed component affords the agent a constant ex post information rent, which weakens the ex ante incentives for effort. We then establish an equivalence for the agent’s payoff set and action choice, between a direct revelation mechanism and the general mechanism. For the principal a truthtelling mechanism strictly dominates because it shields the agent from variations in the ex post payoffs. The Revelation Principle is validated.

1 Introduction

In standard moral hazard problems the outcome of the agent’s action is observable by the principal and may therefore substitute itself for the non-observability of said action. In this case, a complete contract may be designed and conditioned on the outcome. In this paper, attention is paid to the case where the outcome realisation is not observable by the principal either. Thus the information is said to be soft in that it is subject to manipulation. As the current bankruptcy proceedings of the investment bank Lehman Brothers demonstrate, there is no reason to believe that informed agents will truthfully reveal unobserved outcomes. In this sequence of events, the principal is exposed to

*School of Economics, UNSW. Email: g.roger@unsw.edu.au
ex ante moral hazard and also faces a problem of adverse selection ex post. For example, after hiring the CEO, a board of directors may ask of him to report on his actions while on the job. Or a regulated firm may be asked to reveal its production cost after investing in an uncertain technology. This idea may also be extended to the taxation problem, where the agent may undertake some investment to enhance her productivity, which she is required to reveal to the tax authority.

The model resembles that of a standard moral hazard framework, bar for the role of soft information, which introduces a specific, and widely observed, structure of the incentive contract – namely, a two-part tariff. This feature is our first substantive result, beyond the characterisation of the optimal incentive scheme. In the standard moral hazard model (Rogerson (1985), Jewitt (1988), Conlon (2009)), the only incentive problem is that of inducing the principal’s desired action. Where effort monotonically governs the distribution of outcomes, implementing the principal’s preferred action is achieved by conditioning the agent’s compensation (monotonically) on that outcome. Here the principal must also elicit information ex post, and the agent has incentives to systematically distort it upwards. If the compensation is increasing in the private information, who would want to admit to a bad outcome? Consequently, (under a Direct Revelation Mechanism – now DRM) the principal must commit to a state-independent lump-sum payment to the agent at the lower bound – a two-part (but not affine) tariff. This introduces an additional cost to induce effort: it weakens the ex ante effort incentives as it constitutes an insurance against failure, hence the optimal action is further distorted from the standard second-best. The optimal contract must accommodate a fundamental tension between ex post information revelation, which requires a constant transfer, and ex ante effort provision, which is best addressed with a state-contingent compensation. This however differs from the problem of ex post expropriation, which destroys ex ante incentives for effort. Here the agent’s risk aversion combined with the lump sum payment render effort more expensive, so the principal optimally curtails it. We refer to this contract as “conditionally second-best”, in the spirit of Laffont and Tirole (1986). Importantly, (i) it is a low(er)-powered incentive scheme and (ii) its two-part nature emerges endogenously in response to an informational problem, not to satisfy an exogenous participation constraint or by assumption on the part of the modeler. For practitioners or policy-makers, these results suggest that uniformly increasing the steepness of incentive schemes may be counterproductive when information is soft.

1Effort distortion is the real welfare cost in this model.
The second set of results pertains to methodology. A truthful DRM can be characterised. This however is not sufficient to speak to the validity of the Revelation Principle in this context. Indeed, because the lump-sum payment required for truthtelling in a DRM weakens the \textit{ex ante} effort incentives, the existence of the DRM does not imply that the principal wants to insist on costly truthtelling \textit{ex post} – that is, on adopting a DRM in the first place. In technical parlance, it is not obvious that the DRM implements all (or any) equilibrium allocations that the principal could obtain through an alternative mechanism. Still, we can \((i)\) show that an alternative mechanism with an arbitrary message space \(M\) is equivalent in terms of payoffs and effort for the agent, \((ii)\) show that for any action, the DRM offers higher payoffs to the principal in any state and \((iii)\) show that the principal can always implement a higher action under a DRM. Therefore a DRM is a superior mechanism. The reason is that a DRM shields the agent from the lottery over \textit{ex post} payoffs that the general mechanism must generate. So the principal’s cost of providing full insurance decreases. Thus, it is not the truthtelling requirement of the DRM that is costly to the principal, but the imperfect \textit{ex post} observability of outcomes. Since its first incarnations (Green and Laffont (1977), Dasgupta, Hammond and Maskin (1979), Myerson (1979, 1981)), the Revelation Principle has been extended to suit the environment under scrutiny (e.g. Mookherjee and Reichelstein (1990), Bester and Strauz (2001), Martimort and Stole (2002), Kartasheva (2006)). In each of these cases however, not only is the private information \textit{ex post} publicly observable, it is free of \textit{ex ante} interference through the moral hazard problem.

The works closest to ours are Gromb and Martimort (2007), Green and Laffont (1986), Levitt and Snyder (1997) and Diamond (1984). The present model departs from all by adopting soft information in a very strong sense: the principal never observes any outcome.\(^2\) 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 they do not exert any influence on it. In contrast, here a share of the agent’s compensation must be made state independent to induce information revelation. Levitt and Snyder (1997) develop a contracting model in which the agent receives an early (soft) signal about the likely success of the

\(^2\)Gromb and Martimort’s expert(s) receive(s) a soft signal, but whether a project is eventually successful is publicly \((\textit{ex post})\) observable.
project, however the eventual outcome is fully observed by the principal, hence contractible. Here, information can only be observed, and reported, by the agent – and may be audited *ex post* by the principal at some cost. The audit restores partial observability, and is therefore essential as compared to models of costly state-verification (e.g. Khalil (1997)), where it only assists in relaxing the incentive constraint of the agent. Cremer, Khalil and Rochet (1998a,b) allow for an agent to gather costly information about her own type before contracting. To emphasize the point, in these papers that information is still *exogenously* given although *ex ante* unknown to the agent. Here the private information emerges endogenously given although *ex ante* unknown to the agent. Here the private information emerges endogenously. 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, which varies with the true state in a publicly known fashion. $M(.)$-implementable mechanisms exist and need not elicit truthtelling. Their *exogenous* restriction to the set $M(\theta)$ is achieved here *endogenously* through the introduction of the stochastic audit (i.e. there exists an optimal deviation $\tilde{m}(\theta)$ for each type, and the set $\tilde{M}$ is the image of the type space $\Theta$). Diamond (1984) lets an entrepreneur borrow from some lender(s) and alone observe the realisation $\tilde{y}$ of the stochastic outcome. However, lenders can delegate the costly monitoring activity to a bank, which saves them having to elicit the information from the agent. Malcolmson (2009) studies a problem where, as in Gromb and Martimort (2007), the agent acquires soft information and the return to the principal is publicly observable. That soft information may be used by the agent to make a decision yielding the verifiable outcome. The principal may have incentives to distort the decision rule away from the first-best to foster information acquisition.

Next we introduce the model; then the case is made for the use of *ex post* audits. Section 5 characterises the optimal contract presuming of the validity of the DRM. This is confirmed in Section 6 under some conditions. Some of the technical material is relegated to the Appendix.

## 2 Model

A principal delegates a task to an agent. At some cost $c(a)$ increasing and convex, the agent undertakes an action $a \in \mathbb{R}_+$ that yields a stochastic outcome $\theta \in [\underline{\theta}, \overline{\theta}] \equiv \Theta$ with conditional distribution $F(\theta|a)$ and corresponding density $f(\theta|a) > 0$. A higher action is better in the sense

---

3Here, without audit it would be impossible to specify a non-trivial incentive-compatible contract at the information revelation stage, and therefore impossible to implement any effort in the first place.
of first-order stochastic dominance: $F(\theta|a) < F(\theta|a')$ when $a' > a$. The agent alone observes the outcome $\theta$ and reports a message $m$ to the principal, whereupon she receives a transfer $t$. Because she is wealth-constrained, the limited liability constraint $t \geq 0$ applies throughout, and will therefore no longer be mentioned. Her net utility is given by $u(t, a) = v(t) - c(a)$, where $v(t)$ is a completely monotonic, increasing, concave function of the transfer $t$. The principal receives a net payoff $S(t; \theta) = \pi(\theta) - t$, with $\pi' > 0, \pi'' < 0$. The timing is almost standard and corresponds to a complete contracting problem in which the principal is able to commit:

1. The principal offers a contract $\tilde{C} = \langle a, \tilde{D} \rangle$ consisting of an action $a$ and a revelation mechanism $\tilde{D} = (\mathcal{M}, t)$ made of a message space and a transfer
2. The agent accept 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 $m \in \mathcal{M}$
5. Transfers are implemented and payoffs are realised.

$\mathcal{M}$ is a compact set and possesses an order. Under a DRM, $\mathcal{M} = \Theta$ of course. If the true state $\theta$ were observable by the principal, this construct would be a moot point in that it would collapse to the textbook moral hazard problem. Rather, as described, the game is one of cheap talk with commitment, which differs from Crawford and Sobel (1982) where the receiver does not commit to a decision rule $ex \ ante$, but responds simultaneously to the sender’s costless message.

3 The need for audits

For some action $a$ and an outcome $\theta \in \Theta \subset \mathbb{R}$, the agent receives a transfer $t$ from the principal. At the stage of information revelation, effort is sunk so all that matters is the utility $v(t)$ from the transfer $t(m)$. Given the monotonicity of $v(t)$, for a function $t(.)$ increasing, the agent selects some message $\overline{m}$ such that $t(\overline{m}) = \max t$, for $t(.)$ decreasing, some message $\underline{m}$ such that $t(\underline{m}) = \max t$ as

---

$4$Complete monotonicity is characterised by the property $(-1)^{k-1} \frac{d^k v(t)}{dt^k} \geq 0 \forall k \geq 1$, where $k$ indicates the $k$-th derivative.

$5$Even if the state $\theta$ were not verifiable, a Maskin game of state-revelation can elicit truth-telling as a Nash equilibrium of the revelation game.
well, and for \( t(.) \) constant the message can be anything. In the latter case we may want to break indifference in favour of truthful revelation, but there is no good reason to do that. In fact, given that the object is to elicit “effort”, constant transfers guarantee zero effort. Hence either all types pool to the same message or we have a poor mechanism. Auditing restores a measure of ex post observability. It is costly and therefore run with some probability \( \alpha \), for which the principal pays \( k(\alpha) \), increasing and convex. This audit is imperfect and uncovers a lie with probability \( p(m - \theta) \), where \( p : \mathbb{R} \mapsto [0, 1] \) is an increasing \( C^1 \) function. Then the expected utility function of an agent at the revelation stage is

\[
U = v(t(m)) [1 - \alpha(m)] + \alpha(m) p(m - \theta) 0 + (1 - p(m - \theta))v(t(m))
\]

and we have \( \frac{\partial U}{\partial \alpha} = v'[1 - \alpha(m)p] \geq 0 \); \( \frac{\partial^2 U}{\partial \alpha \partial p} = v' \alpha p' \). This is a sorting condition on the expected utility of the agent, akin to the Spence Mirrlees condition. The contract offered in the first stage is no longer \( \tilde{C} \) but \( C = \langle a, D = (\mathcal{M}, t, \alpha) \rangle \).

4 Formulation of the problem

Two problems arise when trying to formulate this problem. First, imposing a constraint on transfers is necessary to address both the information revelation problem and the moral hazard problem. Second, the agent’s ex post message (truthful or not), enters the moral hazard problem, instead of the actual observation of outputs in standard problems. We address these in turn.

4.1 Constraint on transfers

Fix a mechanism \( D \) and consider the agent’s problem after the action \( a \) has been taken; it is sunk at this stage, so she seek a message \( \hat{m}(\theta) \) such that \( v(t(\hat{m}(\theta)))) [1 - \alpha(\hat{m}(\theta))p] \geq v(t(m(\theta))) [1 - \alpha(m(\theta))p] \).

That is, she solves \( \max_{m \in \mathcal{M}} v(t(m)) [1 - \alpha(m)p] \). Her best reply is a message \( \hat{m} \) solving the following first order-condition:

\[
v' t' (1 - \alpha p) - v[\alpha' p + \alpha p'] = 0 \quad (4.1)
\]

6The formulation \( p(m - \theta) \) is necessary to render the agent’s expected utility responsive to the true state \( \theta \). An alternative avenue would be to let \( p = p(\theta) \) and call \( p(.) \) a probability of lying.

7This differentiable approach is shown to be valid in the Appendix.
For a mechanism to be truthful, \( v(t(\theta)) \geq v(t(m)) \left[ 1 - \alpha(m)p \right] \), that is, truth-telling corresponds to a maximum: \( v(t(\theta)) = \max_{m \in \Theta} v(t(m)) \left[ 1 - \alpha(m)p \right] \). Using (4.1), this implies

\[
v' t'(\theta) = v\alpha p'.
\]

But since \( p(.) \) is \( C^1 \), it is symmetric around 0, so \( p'(0) = 0 \) and therefore \( t'(\theta) = 0 \) necessarily. That is, absent further constraints on the transfer \( t(m) \), a truth-telling mechanism calls for transfer that are constant in the message. But \( \text{ex post} \) truth-telling is not the principal’s only concern; he also want to induce some effort, which a constant transfer clearly prevents. Thus we need an additional constraint

\[
t'(m) > 0 \tag{4.2}
\]

which immediately implies that only upward deviations \( (\hat{m} > \theta) \) are of concern.

### 4.2 Agent’s messages

Because the principal’s payoff depends on the true state but the agent’s depends on the message she sends, both \( \theta \) and \( \hat{m} \) figure in the principal’s optimisation. As long as \( \hat{m} = \theta \) (truth-telling) this is a moot point, but of course truth-telling cannot be imposed on the program. That is, we need to account for all possible messages from the agent, not just truthful ones. This is problematic in that we do not possess a distribution over the message space \( \mathcal{M} \), which is left to be arbitrary. However we can restrict our attention to a subset of \( \mathcal{M} \), namely to those messages \( \hat{m} \) satisfying

\[
\hat{m} \in \arg \max_{\mathcal{M}} U(t, \alpha; m, \theta) \tag{4.3}
\]

that is, to optimal messages. Denote this set \( \mathcal{M} \); it is immediate that any truthful message is an element of \( \mathcal{M} \). In the Appendix (Section 8.1), it is shown that \( \mathcal{M} \) is a homeomorphism of \( \Theta \), so it possesses the same structure. Hence they admit the same distribution \( F(\theta|a) \). This result relies on the monotonicity of the optimal message \( \hat{m}(\theta) \), which obtains from the responsiveness of the agent’s expected utility to the true state \( \theta \).

---

8Should the first-order condition fail, then \( \hat{m} = \overline{\theta} \) since \( v' t'(1 - \alpha) - v\alpha p' > 0 \) if \( t' > 0 \) and \( \hat{m} = \underline{\theta} \) otherwise. This is not relevant to the characterisation of a DRM, as soon will be more evident. Please see Lemma 6 in the Appendix for further derivations.
4.3 Statement of the problem

With this we can state the principal’s problem wanting to implement some action $a$ and seeking some ex post report $m$ as

**Problem 1**

$$\max_{a,t,\alpha} \int_\Theta \pi(x) - t(1 - \alpha(\hat{m}(x)))p|dF(x|a) - k(\alpha)$$

s.t. (4.2), (4.3) and

$$\int_\Theta U(t, \alpha; \hat{m}(x))dF(x|a) - c(a) \geq 0 \quad (4.4)$$

$$\int_\Theta U(t, \alpha; \hat{m}(x))dF(x|a) - c(a) \geq \int_\Theta U(t, \alpha; \hat{m}(x))dF(x|\tilde{a}) - c(\tilde{a}) \quad (4.5)$$

where the last two constraints are the standard participation and moral hazard constraints. The novelty of this paper the introduction of the information-revelation constraint (4.3). It is not an ex ante constraint as the agent knows the state $\theta$ at the time of revelation. No interim participation constraint is required as it is addressed by the ex ante condition (4.4) – the agent has already committed to the mechanism.

5 Optimal Contract

In attempting to solve the ex ante Problem 1 while inducing truthtelling, it is neither valid nor necessary to turn (4.3) into a truthtelling constraint of the principal’s program by letting $M = \Theta$. It is not valid because, as $\hat{m}(\theta) > \theta \ \forall \theta$ (Lemma 5, Appendix), for any pair $\alpha, t$ there always exists some value $\theta'(\alpha, t) < \bar{\theta}$ such that $\hat{m}(\theta) = \bar{\theta} \ \forall \theta \geq \theta'(\alpha, t)$ – see the Appendix, Section 8.1. That is, for states above $\theta'(\alpha, t)$, the agent can only report at most $\theta$. But then the message space becomes a truncation of $\hat{M}$ and no longer a homeomorphism of $\Theta$, which invalidates the formulation of Problem 1. And it is not necessary because the problems of providing effort incentives (ex ante) and information revelation (ex post) are in fact separable.

5.1 Ex post information revelation under a DRM

Because auditing restores a measure of ex post observability we can characterise a direct mechanism. In the Appendix (Section 8.1) we formalise this statement. Existence of a DRM is not sufficient
to validate the Revelation Principle in the present setting, in that this construction alone does not guarantee that the optimal allocation is implementable through a DRM. We postpone the validation of this approach until after the characterisation of the optimal DRM (Section 6).

As we know from subsection 4.1, letting $M = \Theta$, the first-order condition of the agent’s problem immediately implies $v't'(\theta) = 0$, whence the transfer sufficient for truth-telling is a constant. But of course this constant cannot be arbitrary. In a standard problem of adverse selection, a boundary condition at $\theta$ is given by the (interim) participation condition: $U(\theta) \geq U_0$ – whatever this (exogenous) outside option may be. In this problem, participation has already been committed to by the time we reach the reporting stage, so there is no such condition. Here the boundary condition is not only necessary to characterise the transfer required to induce truthful revelation, it will later be shown it is also sufficient. At the boundary, the agent must be made indifferent between reporting her private information and an alternative message $\hat{m}$ at $\theta$, i.e. $U(\theta) \geq v(t(\hat{m})) [1 - \alpha(\hat{m})p] > 0$, with the second inequality following from Lemma 5 (Appendix). Therefore the least-cost boundary that is necessary for truth-telling at $\theta$ is a function $\bar{U}(t, \alpha; \theta)$ such that

$$\bar{U}(t, \alpha; \theta) \equiv v(t(\hat{m})) [1 - \alpha(\hat{m})p]$$

(5.1)

where $\hat{m}$ solves (4.3) at $\theta$. Using the envelope theorem, $\frac{\partial U}{\partial t} = v'(1 - \alpha p) > 0$ and $\frac{\partial U}{\partial \alpha} = -vp < 0$. These results are summarised in our first Proposition. Let $h(\bar{U})$ be the monetary equivalent of $\bar{U}$.

**Proposition 1** A truth-telling direct revelation mechanism exists. The compensation schedule must take the form of a two-part tariff denoted

$$t(\theta) = \tau(\theta) + h(\bar{U}), \quad \theta \in \Theta,$$

with $\tau(\theta) = 0$, $\tau' \geq 0$ and $\bar{U}$ defined by (5.1).

Here the boundary condition (5.1) is a lump sum payment (given some schedules $t, \alpha$). While the first term of the tariff $t(\theta)$ is state-independent, the variable component $\tau(\theta)$ of the agent’s compensation is to be used to provide incentives for effort. Thus Proposition 1 imposes some structure on the payment schedule beyond what is known from standard models. The formal claim of existence can be found in the Appendix (Theorem 1), and follows directly from Lemmata 4 to 6.

---

9Either the first-order condition binds or $\hat{m} = \theta$ and the message can no longer increase, so $\hat{m} = \theta$ is a maximum.
and Proposition 7 (also in the Appendix). Under the two-part tariff \( h(U) + \tau \), the agent’s \textit{ex post} information rent can be defined as \( V(U, \theta) \equiv v(h(U) + \tau(\theta)) - v(t(\theta)) \), where \( t(\theta) \) is the schedule that solves the standard moral hazard problem under full observability.

### 5.2 Optimal compensation

We proceed in two steps, first seeking to minimise the \textit{ex post} rent to the agent by choice of an optimal audit probability, and second selecting the optimal transfer \( \tau(\cdot) \) to provide her with \textit{ex ante} incentives. A fully committed principal seeks to minimise the expected rent of an agent:

\[
\int \mathcal{P} V(U, \theta) dF(y|a),
\]

which is equivalent to minimising the sufficient level of \textit{ex post} utility required for truthtelling: \( U(t, \alpha; \theta) \). Then \( \alpha^* \) is defined as

\[
\alpha^* \in \arg \min_{\alpha \in [0,1]} U + k(\alpha).
\]

with first order condition

\[
k'(\hat{m}(\theta)) = v(t(\hat{m}(\theta))) p(\hat{m}(\theta) - \theta)
\]

and since \( \frac{d\hat{m}(\theta)}{d\alpha} < 0 \) (Lemma 5, Appendix), this first-order condition is sufficient to identify a minimiser. The optimal audit probability strictly limits the rent; as we know from the Appendix (Section 8.1), rent and audit probability are substitutes. With this in hand we can turn to the optimal transfer \( \tau(\theta) \). Adopting the first-order approach (see Jewitt (1988), Araujo and Moreira (2001) or Conlon (2009) for multiple validations), Constraint (4.5) is turned into a first order condition with respect to the action \( a \) chosen by the agent. Presuming of truthtelling for now – that is, presuming Condition (5.1) is sufficient, which will be verified later – the principal solves

**Problem 2**

\[
\max_{a,\tau} \int \Theta \pi(x) - [h(U) + \tau] dF(x|a)
\]

s.t. (5.1) and

\[
\int \Theta v(h(U) + \tau) dF(x|a) \geq c(a)
\]

\[
\int \Theta v(h(U) + \tau) dF_a(x|a) = c'(a)
\]

where the optimal audit probability \( \alpha^* \) is substituted in (5.1). Problem 2 is a standard formulation up to that very constraint – the lump-sum payment. Under Constraint (5.1), the agent reports
truthfully for any transfer. Once the solution $\tau^*$ is characterised, $\mathcal{U}(t, \alpha^*; \theta)$ is determined at $\theta$ by (5.1), which defines a fixed point.

**Proposition 2** The principal offers a “conditionally second-best” contract made up of an increasing, concave compensation schedule $\tau^*(\theta) + h(\mathcal{U})$ satisfying

- an audit probability $\alpha^* > 0$ given by (5.2);
- the standard first-order condition $\frac{1}{\theta} = \lambda \frac{f_p}{\theta} + \mu$ of Problem 2; and
- the fixed point $\mathcal{U} = (1 - \alpha^* p)v(t(m(\theta)))$.

The first-order condition is completely standard; observe in particular that no ex ante rent is left to the agent: the participation constraint is saturated. The familiar condition $\frac{1}{\theta} = \lambda \frac{f_p}{\theta} + \mu$ arises because conditional on truthtelling, the problem is simply one of moral hazard. With this results we also can claim

**Proposition 3** The agent receives a constant ex post information rent $V(\mathcal{U}, \theta)$, hence Condition (5.1) is sufficient for truthtelling for all types $\theta \in \Theta$.

### 5.3 Optimal action choice

We are now in a position to complete the characterisation of the moral hazard problem under the DRM, which has become almost standard. However, while the substitutability of $h(\mathcal{U})$ and $\tau$ has no direct effect the principal’s objective function, it implies less powerful incentives for the agent – lower marginal returns on effort. This is our second main result.

**Proposition 4** Imperfect observability undermines the provision of effort. Action $a^*$ solving (5.4) is lower than in the standard moral hazard problem with perfect outcome observability.

The truthtelling constraint imposes a non-trivial fixed component $h(\mathcal{U})$ as part of the agent’s compensation. This guarantees a fixed payment regardless of the state she reports, including a positive transfer in the worst state. As a result, the ex ante incentives of the agent are softened. The effect of this guaranteed payment resembles that of an outside option, although it is not on two grounds. First, if rejecting the contract the agent receives nothing and second, it is determined endogenously at the interim stage – after the contract has been accepted. A direct consequence of Proposition 4 is our next result, which completes the characterisation of the contract $\mathcal{C}$. 
Proposition 5 The principal selects a “conditionally second-best” action $a^{SB}$ solving

$$\int_{\Theta} \pi(\theta) - [h(U) + \tau]dF_a(\theta| a^{SB}) + \lambda \left( \int_{\Theta} v(h(U) + \tau)dF_{aa}(\theta| a^{SB}) - c''(a^{SB}) \right) = 0 \quad (5.5)$$

which is lower than he would in the standard moral hazard problem with perfect outcome observability. That is, a low-powered incentive contract is offered to the agent.

This is reminiscent of Laffont and Tirole (1986), where there is no first-order (direct) effort distortion but one that is only indirect (through the revelation problem). In their paper, type and effort are substitutes (but types are ex ante, hard information) and revelation is obtained through ex post cost observability. In contrast here truth telling is achieved by a combination of the optimal lump sum $U$ and the audit probability.

6 Validity of the direct revelation mechanism: the Revelation Principle

It is now high time to enquire as to whether an optimal mechanism for the principal should include a truth telling DRM. The reason is that the ex post information revelation problem requiring the compensation $\tau + U$ under is in tension with the ex ante effort incentives in that the principal induces a lower action $a^{SB}$. So it is not obvious yet that the principal could not be better off under an alternative mechanism, under which he may be able to elicit more effort. Relying on a DRM turns out, perhaps surprisingly at first sight, to make the principal strictly better off. We proceed in two steps, first establishing an equivalence result for the agent, and then showing that the principal’s ex ante choice of action must be higher under the DRM. Two definitions need now be introduced. Let $T$ be the space of allocation for a generic mechanism. Retain the notation $M$ and let $d: M \rightarrow T$ be a generic allocation rule.

**Definition 1** Let $u(.)$ be a generic payoff function. Two mechanisms $(M, d(m)), (M', d'(m'))$ are payoff-equivalent when $u(d(m)) = u(d'(m'))$.

**Definition 2** Two contracts $(a, (M, d(m)), (a', M', d'(m)))$ are effort-equivalent when $a = a'$, where $a'$ is the optimal action (i.e. solves (5.5)) under $(M', d'(m))$.  

12
6.1 An equivalence result for the agent

Calling on Definition 1 we can now claim:

**Lemma 1** Fix an action $a$. The general mechanism $\langle \mathcal{M}, \alpha^*(m), t^*(m) \rangle$ and the direct mechanism $\langle \Theta, \alpha^*(\theta), U^*, \tau(\theta) \rangle$ are payoff-equivalent for the agent.

In some sense this is a re-statement of the incentive compatibility requirement, which has to be satisfied by construction of the truthtelling mechanism. Indeed, recall that condition (5.1) prescribes that the agent be paid exactly her optimal deviation at $\theta$. Since that condition is also sufficient, the *ex post* utility of the agent varies at the same rate across the state space under either mechanism. So her payoff is the same in either case.

**Lemma 2** Take the general mechanism $\langle \mathcal{M}, \alpha^*(m), t^*(m) \rangle$ and the DRM $\langle \Theta, \alpha^*(\theta), U^*, \tau(\theta) \rangle$ such that the truthtelling condition (5.1) holds. These mechanisms are effort equivalent for the agent.

Thus faced with a contract that yields the same expected utility, the agent takes the same action; that is, the moral hazard constraints are equivalent. Although the agent’s payoffs are equivalent under either mechanism given some action $a$ fixed, said action needs not be the same under different mechanism because the principal steers it. These equivalence claims are silent as to the principal’s choice of action $a$, and as such cannot yet be read as a statement about the Revelation Principle. In particular, it does not imply that the principal wants to induce the same action in either case. We turn now to this question.

6.2 Principal’s choice of a contract

First we show that Lemma 1 has no equivalent for the principal, for which we need to alter Definition 1. This is of relevance to understand the principal’s incentives to choose an optimal action.

**Definition 3** Let $\pi(.)$ be a generic payoff function. A mechanism $\langle \mathcal{M}, d(m) \rangle$ is payoff-dominant over an alternative $\langle \mathcal{M}', d'(m) \rangle$ when $\pi(d(m)) \geq \pi(d'(m'))$.

**Lemma 3** Fix an action $a$. The DRM is strictly payoff-dominant for the principal.

While the agent is indifferent between either mechanism (Lemma 1), the risk neutral principal can economise on the transfer he offers. In steering the agent to truthful revelation by offering her
a DRM, the principal shields the agent from variation in the *ex post* transfer. Consequently the total transfer $h(U) + \tau$ necessary to make the agent indifferent between the outcomes under the truthful mechanism and the general mechanism is lower than $t(\tilde{m})$.\textsuperscript{10} Thus selecting a mechanism implies choosing from a non-trivial set of optimal actions to implement. A sufficient condition for the principal to choose the truthtelling mechanism is that his preferred action under the DRM be no lower than under the general mechanism. Finally we have:

**Proposition 6** Let $a'$ be the principal’s preferred action under the general mechanism and $a^{SB}$ solve (5.5). $a^{SB} > a'$.

Because the principal economises on the expected transfer paid to the agent under a truthful mechanism, he is also willing to demand a higher costly action. Together with Lemma 3, this directly implies that the principal is made better off.

### 6.3 Discussion

Proposition 4 may lead the analyst to conclude that the truthtelling requirement of the DRM is too costly, and to conjecture that there may be less onerous (non-truthful) mechanisms to avail. But it is poor *ex post* observability, rather than truthtelling, that is the source of the effort distortion. Indeed, up to the choice of action $a$ by the principal, Lemma 1 tells us that the payoff to the agent is the same under a DRM as any other mechanism. This owes to poor *ex post* information, which imposes either a non-trivial fixed component $h(U)$ as part of the agent’s compensation, or exposes the principal to a lie – both yielding the same expected payoff to the agent, for the same action. This is Condition (5.1). Both equally soften the incentives as the agent is guaranteed either a fixed payment (under the DRM), or (optimally) overstates her private information $\tilde{m}(\theta)$ under the general mechanism and guarantees herself a positive expected transfer – this is Lemma 2. Thus the rent arises from the lack of *ex post* observability, which is only imperfectly restored by the audit mechanism.

From the risk-neutral principal’s standpoint, insisting on truthtelling guarantees a higher *ex post* payoff in any state, because he can save on the insurance cost. In the alternative, the agent is exposed to a lottery over *ex post* payoffs, which requires a higher transfer from the principal (in

\textsuperscript{10}The cost of fully insuring the agent decreases under the direct mechanism.
any state of the world). The benefit however is two-fold: the principal enjoys a direct increase in his \textit{ex post} payoff from a lower transfer, and an increase in the action that can be required of the agent. In equilibrium, under a truthtelling mechanism, the agent works harder and the principal enjoys higher expected returns than if allowing the agent to misreport her information – however not than under complete observability.

6.4 \textbf{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 $M(.)$ is exogenous and publicly known. They 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 our model, although it corresponds to a game of “unidirectional distortions with an ordered space” (to use their words) – example a(2) in Green and Laffont). Because the agent has a unique optimal deviation for each type, the set $M(\theta) = \{ \bar{\theta} | \bar{\theta} \in \Theta, \bar{\theta} \geq \theta \}$ (to use their notation) collapses to a singleton for each type, whence no nesting condition can possibly hold. In addition, $\bar{M} \subseteq \Theta$. Instead, truthtelling is restored through the two-part tariff (and the audit).

7 \textbf{Conclusion}

When a principal cannot observe the outcome of his agent’s action in a moral hazard framework and needs to elicit this information from that very agent, he faces a problem of \textit{ex post} adverse selection as well. This is costly to the principal in two ways: first, it requires paying a rent to the agent to obtain any truthful information. Second, it distorts the \textit{ex ante} effort incentives further away from the standard second best because the provision of an \textit{ex post} rent reduces the marginal return on effort. The source of these distortions is not the truthtelling requirement imposed by the Direct Revelation Mechanism. In fact, not only can the Revelation Principle be applied, a truthtelling mechanism affords the principal strictly higher payoffs, both \textit{ex post} because the principal saves on costly insurance, and \textit{ex ante} because he can induces a higher action on the part of the agent. The optimal contract presents two feature, which may be perceived as weaknesses: it is low-powered

\footnote{When $\bar{M} \subseteq \Theta$ for some exogenous reason, say, some pooling occurs at $\bar{\theta}$ – see Lemma 6.}
and requires a lump-sum payment in the case of failure. On the other hand, it may be argued that widely observed, high-power contracts (as those of CEOs) may be excessively steep and not conducive of truthful revelation. A dynamic extension is an obvious avenue to attempt to address these shortcomings, and possibly reconcile these two perspectives.

8 Appendix

8.1 Characterisation of the optimal message space under a general mechanism and feasibility of DRMs

We characterise the set \( \mathcal{M} \) of equilibrium messages when the mechanism is not restricted to be a DRM. It is important for two reasons. First, the problem can be formulated using the underlying distribution of the state \( \theta \). Second, it is useful in comparing direct and general mechanisms. To proceed, relax a DRM’s requirement that \( m \in \Theta \). Let \( m \in \mathcal{M} \) where \( \mathcal{M} \) is an arbitrary set endowed with an order. Define the transfer and audit probability as \( t(m) : \mathcal{M} \mapsto \mathbb{R} \) and \( \alpha(m) : \mathcal{M} \mapsto [0, 1] \).

At the revelation stage, an agent selects a message \( \hat{m} \) solving \( \max_{m \in \mathcal{M}} v(t(m))[1 - \alpha(m)p] \) with necessarily binding first-order condition (4.1) – as \( \mathcal{M} \) may be made sufficiently large.

Claim 1 Condition (4.1) admits a unique maximiser.

Proof: Directly from the sorting condition \( \frac{\partial^2 U}{\partial m^2} = v' \alpha' p \), which displays a constant sign. ■

Hence there exists an optimal message for each type. Let \( \hat{m}(t, \alpha; \theta) \) be the solution to equation (4.1). For any transfer function \( t(m) \) and probability of audit \( \alpha(m) \), it can be shown that \( \hat{m}(t, \alpha; \theta) \) varies continuously in \( \theta \) (Theorem of the Maximum, Berge (1963)). That is, the agent’s reporting behaviour generates a bounded set \( \hat{\mathcal{M}} = \{m \in \mathcal{M} | m \in \arg\max U(t, \alpha; \theta), \theta \in \Theta\} \). Else, \( \hat{m} = \overline{m} \) for some (arbitrary) upper bound \( \overline{m} < \hat{m}(t, \alpha; \overline{\theta}) \) of \( \mathcal{M} \) if

\[
\frac{v'(t')}{v(t(m))} \geq \frac{\alpha' p}{1 - \alpha(m)p}
\]  

where any derivative with respect to \( m \) is understood to be a left-hand derivative.

Lemma 4

\[
\frac{d \hat{m}}{d \theta} > 0
\]
Proof: Continuity of the solution $\hat{m}(t, \alpha; \theta)$ follows from the Theorem of the Maximum. Monotonicity can be easily established; jumping ahead and calling on the Theorem of Lebesgue, $\hat{m}(t, \alpha; \theta)$ is a.e. differentiable, except at most for a finite set of points. Differentiate with respect to $\theta$ and rearrange.

For notational convenience, let $\min \hat{m} \in \hat{M} = \hat{m}(\theta)$. We also note

**Lemma 5** For any (weakly increasing) transfer function $t(m)$ and audit probability $\alpha(m)$,

$$\forall \theta \in \Theta, \ \hat{m}(\theta) > \theta.$$  

Furthermore, $\frac{d\hat{m}(\theta)}{dt} < 0$ and $\frac{d\hat{m}(\theta)}{d\alpha} < 0$ so that $\frac{d\hat{m}(\theta)}{d\alpha} < 0$ as well.

Proof: Notice first that in the standard moral hazard problem, the first-order condition $\frac{1}{\nu} = \mu + \lambda \frac{L}{T}$ immediately implies that $t(\theta) > 0$. Now take $t(\theta)$, then there exists some $\hat{m}(\theta)$ such that $v(t(\theta)) < v(t(\hat{m}(\theta)))(1 - \alpha p)$. So $\hat{m}(t, \alpha; \theta) > \theta$ necessarily. To see why, take the first-order condition $v'(\hat{m}) (1 - \alpha p) - v(\alpha') = 0$ at $\theta$ and let $\hat{m} \to \theta$. Then the LHS tends to $v' > 0$: there is always an incentive to increase the message $\hat{m}(\theta)$ starting from $\hat{m}(\theta) = \theta$. That is, for any $\theta \geq \theta$, there exists some $\hat{m}(\theta)$ such that $v(t(\theta)) \leq v(t(\hat{m}(\theta)))(1 - \alpha p)$. Since $(1 - \alpha p) < 1$, we necessarily must have $\hat{m}(\theta) > \theta$. For the second set of statements, continuity and monotonicity can be easily shown, so that $\hat{m}$ is a.e. differentiable with respect to $t$ and $\alpha$. Fix an arbitrary type such that the first-order condition (4.1) holds, differentiate it with respect to each of the variables and re-arrange. Using the fact that the SOC holds the claims follow. The results extend all the way to $\theta = \theta$ from the right-hand side by (right-hand) continuity.

Lemma 5 establishes that the agent’s optimal reporting choice systematically overstates the truth in all states $\theta$. In addition, it tells us that there is less “over-reporting” when the transfer increases and when the probability of audit increases. The latter claim is quite obvious, the first one less so: increasing $t(m)$ renders lying more costly as there is more to lose for any actual state realisation. The transfer and the audit act as substitutes in the agent’s information revelation problem. We now complete the description of the set of optimal messages $\hat{M}$, with $\max \hat{m} \in \hat{M} = \hat{m}(\theta)$.

**Lemma 6** Take an arbitrary message space $\mathcal{M}$ with an arbitrary upper bound $\mathcal{m}$, and any $t(m), \alpha(m)$, then

$$\hat{m}(\theta) = \min \{\mathcal{m}, \hat{m}(t, \alpha; \theta)\}$$
where \( \hat{m}(t, \alpha; \vec{\theta}) \) solves Condition (4.1) at \( \vec{\theta} \). If \( \hat{m}(\vec{\theta}) = \overline{m} \), then there exists a threshold \( \theta'(t, \alpha) \leq \vec{\theta} \) such that \( \hat{m}(\theta) = \vec{\theta}, \forall \theta \geq \theta'(t, \alpha) \) with \( \frac{d\theta'(t, \alpha)}{dt} > 0, \frac{d\theta'(t, \alpha)}{d\alpha} > 0, \forall \theta'(t, \alpha) < \vec{\theta} \).

**Proof:** For the first statement, take any increasing monotone transfer function \( t(m) \) and audit probability \( \alpha(m) \) defined on \( \mathcal{M} \). For any type \( \theta \) and any message \( m \neq \theta \), either FOC (8.1) or (4.1) holds. In the former case, the optimal message is exogenously bounded at \( m \) of \( \mathcal{M} \). Then the existence of \( \theta'(t, \alpha) \) follows directly from FOC (8.1) and from Lemma 5. More precisely, at \( \theta'(t, \alpha) \), the agent’s utility \( U = t(m)[1 - \alpha(m)p] \) is not differentiable; but the left-hand derivative is given by (4.1), while the right-hand derivative is zero (since \( \hat{m} = \vec{\theta} \)). When Condition (4.1) is relevant, the maximiser of \( U(\theta) \) is unique (Claim 1), and \( \max U(\theta) = U(\vec{\theta}) \). Therefore \( \hat{m}(t, \alpha; \vec{\theta}) = \max \hat{m}(\theta) \). The second set of statements follows directly from the existence of \( \theta'(t, \alpha) \) and from Lemma 5. Indeed, \( \frac{d\hat{m}}{dt} < 0 \) wherever the FOC (4.1) holds. In particular, the left-hand derivative \( \frac{d\hat{m}(\alpha, \dot{t}, \theta)}{dt}|_{\theta = \theta'(t, \alpha)} < 0 \) for some transfer \( \dot{t} \). Therefore \( \hat{m}(\alpha(t; \theta'(t, \alpha))) < \vec{\theta} \) for any \( t > \dot{t} \). Hence by the first claim, there exists some other \( \theta'(t, \alpha) > \theta'(\dot{t}, \alpha) \) such that \( \hat{m} = \vec{\theta}, \forall \theta \geq \theta'(t, \alpha) \). Since the threshold \( \theta'(t, \alpha) \) is a monotonic function, it is a.e. differentiable except for finite set of points (Theorem of Lebesgue).

As the true state approaches the upper bound of the support, the threat of audit does not have enough bite to prevent deviations all the way to the top. However, since over-reporting is dampened by increases in \( \alpha, t \), the threshold \( \theta'(t, \alpha) \) above which the agent lies without restraint increases. So both the transfer and the audit probability improve information revelation, but at different costs. Collecting these preliminary results, we have

**Proposition 7** Fix \( t(m) \) and \( \alpha(m) \), the optimal reporting set is a compact interval: \( \hat{\mathcal{M}} = [\hat{m}(\theta), \hat{m}(\vec{\theta})] \).

Using this technical result we can finally claim

**Theorem 1** A Direct Revelation Mechanism exists.

**Proof:** Since \( \hat{m}(\theta) \) is unique for each type \( \theta \) card\( \hat{\mathcal{M}} = \text{card}\Theta \). Because \( \hat{m}(\theta) \) is monotonic, it is invertible. That is, the mapping \( m(\theta) \) is a homeomorphism from \( \Theta \) to \( \hat{\mathcal{M}} \). The general mechanism is a pair of functions \( \langle \hat{\alpha}(m(\theta)), \hat{t}(m(\theta)) \rangle \), which are both continuous and monotonic. Define the direct mechanism \( \langle \alpha(\theta), t(\theta) \rangle \) as \( \alpha \equiv \hat{\alpha} \circ m \) and \( t \equiv \hat{t} \circ m \). For any \( \theta \in \Theta \), there exists a unique pair \( \langle \alpha(\theta), t(\theta) \rangle \) as well. ■
8.2 Proofs

Proof of Proposition 1: Take some \( t(m), t'(m) > 0 \) solving a standard moral hazard problem, then by Lemma 5, for any \( \alpha \in (0, 1) \), there exists some \( \hat{m} \) such that \( v(t(\hat{m}(\hat{\theta})))(1 - \alpha p) > v(t(\theta)) \). Therefore construct a compensation schedule of the form \( t(m) = h(U) + \tau(m) \), such that Condition (5.1) holds, and with \( \tau' > 0 \) to satisfy constraint (4.2).

Proof of Proposition 2: Take \( \alpha^* \) as given by (5.2) for some transfer function \( t(m) \). Attaching multipliers \( \mu \) and \( \lambda \) to constraints (5.3) and (5.4), the first-order condition with respect to \( \tau \) reads

\[
\frac{1}{v'} = \lambda \frac{f_a}{f} + \mu, \quad \forall \theta \in \Theta \quad (8.2)
\]

Up to the argument of \( v'(.) \) this is exactly the condition derived by Jewitt (1988), so we know that both \( \lambda \) and \( \mu \) are positive. Observe that \( U = (1 - \alpha p)v(h(U) + \tau(\hat{m})) \) defines a fixed-point problem with a unique solution since the RHS is a contraction. To see why, differentiate with respect to \( U \)

\[
\frac{d}{dU} (1 - \alpha p)v(h(U) + \tau(\hat{m})) = (1 - \alpha p)v' \left( h' + \tau' \frac{d\hat{m}}{dU} \right) - v(.) \left[ \alpha' p + \alpha p' \right] \frac{d\hat{m}}{dU} \]

\[
= (1 - \alpha p)v'h' + \frac{d\hat{m}}{dU} [(1 - \alpha p)v' \tau' - v(.) (\alpha' p + \alpha p')] \]

\[
= (1 - \alpha p) \leq 1
\]

by application of the envelop theorem and where \( h(.) \) is the monetary equivalent of \( U \). To see that \( \tau(.) \) is increasing, simply observe from the first-order condition (8.2) that the RHS is increasing and concave in \( \theta \) (by assumption on \( f_a/f \)), and therefore so is the LHS. That is, \( v'(.) \) is decreasing convex, so \( \tau(.) \) is necessarily increasing by complete monotonicity of \( v(.) \). To show concavity, rewrite the FOC as \( v' - \left( \mu + \lambda \frac{f_a}{f} \right)^{-1} = 0 \) and differentiate w.r.t. \( \theta \) to find \( v'' \tau' + \lambda \frac{d}{d\theta} \frac{f_a}{f} / \left( \mu + \lambda \frac{f_a}{f} \right)^2 = 0 \). This verifies \( \tau' > 0 \). Re-arrange this expression and redefine the variables

\[
\tau' = -\lambda \frac{1}{v''} \left( \mu + \lambda \frac{f_a}{f} \right)^2
\]
Then $\tau'' \geq 0 \iff \left( \frac{dY}{d\theta} X + \frac{dX}{d\theta} Y \right) \leq 0$. Rewrite the second condition as

$$
\begin{align*}
\frac{dY}{d\theta} X & \leq -\frac{dX}{d\theta} Y \\
\frac{d}{d\theta} \ln -Y & \leq \frac{d}{d\theta} \ln X \\
\frac{d}{d\theta} \ln -\frac{1}{v''} & \leq \frac{d}{d\theta} \ln \left( \frac{\frac{d}{d\theta} \left( \frac{f}{a} \right)}{\left( \mu + \lambda \frac{f}{a} \right)^2} \right)
\end{align*}
$$

where the second line obtains because $Y < 0$. Since the ratio $\frac{f}{a}$ is increasing concave, the RHS is negative. It is immediate to verify by differentiation that the LHS is positive, so the necessary and sufficient condition cannot hold. Last, the participation constraint (5.3) of the agent binds, so she received no *ex ante* rent.

**Proof of Proposition 3:** Truth telling for all types requires extending the condition $U = v(t(\hat{\mu}(\theta)))(1 - \alpha(\hat{\theta}))p(\hat{\mu}(\theta) - \theta)$ to any realisation of $\theta$. That is, we need $U(\theta) \geq v(t(\hat{\mu}(\theta)))(1 - \alpha(\hat{\theta}))p(\hat{\mu}(\theta) - \theta)$. Define the agent’s information rent under the two-part tariff as $V(U, \tau) = v(h(U) + \tau(\theta)) - v(t(\theta))$, where $t(\theta)$ solves the standard moral hazard problem. By differentiation

$$
\frac{dV(U, \theta)}{d\theta} = v'(h(U) + \tau) \left[ h' \frac{dU}{d\theta} + \tau' \right] - v'(t)t'
\begin{align*}
&= v'(h(U) + \tau)\tau' - v'(t)t' \\
&= \varphi(s) \left[ \tau'(\theta) - t'(\theta) \right], \ s = h + \tau, t
\end{align*}
$$

where the second line obtains since $U$ is invariant in the private information (it is defined at $\hat{\theta}$) and the third one holds since $v'(h + \tau) = \left( \mu + \lambda \frac{f}{a} \right) = v'(t) \forall \theta$: the FOC are identical for all states $\theta$. From the proof of Proposition 2, $\tau'$ and $t'$ can only vary according as $v''(h + \tau)$ and $v''(t)$, that is, by $\varphi'(s)$. More precisely, using the notation of the proof of Proposition 2,

$$
\begin{align*}
\frac{dV(U, \theta)}{d\theta} &= \varphi(s) \left[ \tau'(\theta) - t'(\theta) \right] \\
&= -\lambda X \varphi(s) \left[ \varphi'(s) - \varphi'(s) \right] \\
&= 0
\end{align*}
$$

since $v'(h + \tau) = v'(t) = \varphi(s) \forall \theta$. Since the rent is constant in the type, the agent has no incentive to distort her message, so (5.1) is also sufficient. ■
Proof of Proposition 4: Since \( t = \tau + h(U) \), \( U > 0 \) to satisfy truthtelling (Condition (5.1)). Take any (total) transfer \( \tilde{t} \) such that \( 0 \leq h(U) < h(U) \); for any action \( a \)

\[
\mathbb{E}[v(\tau + h(U))|a] = \int \mathbb{E}[v|a] \, dF(\theta|a) = \mathbb{E}[v(\tilde{t})|a],
\]

by concavity of \( v(\cdot) \). Since \( \mathbb{E}[v(t)|a] \) is concave in \( a \) under the assumptions of the first-order approach (see Jewitt (1988)),

\[
\int \mathbb{E}[v|a] \, dF(\theta|a) = c'(a^*) < \int \mathbb{E}[v(\tilde{t})|a] = c'(\tilde{a}),
\]

whence \( a^* < \tilde{a} \) (where \( \tilde{a} \) solves (5.4) under \( \tilde{t} \)) since \( c(a) \) is increasing convex.

Proof of Proposition 5: The first-order condition of Problem 2 under the truthtelling mechanism with respect to the action \( a \) is the standard first-order condition given by (5.5) with the participation constraint binding since \( \mu > 0 \). As in Proposition 4, the marginal return to the principal is lower – or, an action \( a^{SB} \) solving (5.5) is more expensive than would be absent \( h(U) \). Hence, given a pair of optimal \( t^*, \alpha^* \), the principal selects a lower action \( a^{SB} \).

Proof of Lemma 1: Fix \( a \), then the distribution \( F(\theta|a) \) is fixed and identical under either revelation mechanism. The lowest transfer \( h(U) \) under the DRM is given by the boundary condition (5.1): \( U = v(t(\hat{\mathcal{m}}(\theta)))(1 - \alpha(\hat{\mathcal{m}}(\theta))p) \). This is exactly the expected utility of the agent at \( \hat{\theta} \) under the general mechanism. By the sufficiency claim of Proposition 1, ex post payoffs must be identical for any realisation of \( \theta \). Given an identical distribution \( F(\theta|a) \), ex ante payoffs are also identical.

Proof of Lemma 2: Take transfers \( t(\hat{\mathcal{m}}) \) and \( h(U) + \tau \) such that (5.1) holds. Then by the sufficiency claim of Proposition 1, \( \forall \theta, \, v(h(U) + \tau(\theta)) = v(t(\hat{\mathcal{m}}(\theta)))(1 - \alpha(\hat{\mathcal{m}}(\theta))p) \). Therefore the agent’s problem \( \max_a \int \mathbb{E}[v(t(\hat{\mathcal{m}}(\theta)))(1 - \alpha^{*}p)dF(x|a) - c(a)] \) yielding moral hazard constraint (5.4) and the problem \( \max_a \int \mathbb{E}[v(t(\hat{\mathcal{m}}(\theta)))](1 - \alpha^{*}p)dF(x|a) - c(a) \) with FOC

\[
\int \mathbb{E}[v(t(\hat{\mathcal{m}}(\theta)))](1 - \alpha^{*}p)dF_a(x|a) = c'(a)
\]

are identical for the agent. Therefore their solutions are identical.

Proof of Lemma 3: From incentive compatibility,

\[
\forall \theta, \, v(h(U) + \tau(\theta)) = v(t(\hat{\mathcal{m}}(\theta)))(1 - \alpha^{*}p) = v(CE)
\]

\footnote{With complete \textit{ex post} observability, the optimal \( U \) is zero.}
where $CE$ stands for certainty equivalent. Then we know
\[
v(h(U) + \tau(\theta)) = v(t(\hat{m}(\theta)))(1 - \alpha^*p) < v(t(\hat{m}(\theta)))(1 - \alpha^*p)
\]
by Jensen’s inequality, whence
\[
\forall \theta, \quad h(U) + \tau(\theta) = g(v(t(\hat{m}(\theta)))(1 - \alpha^*p)) < t(b_m(\theta))(1 - \alpha^*p)
\]
where $g \equiv v^{-1}$ is an increasing function. It then follows that
\[
\int [h(U) + \tau(\theta)]dF(\theta|a) < t(b_m(\theta))(1 - \alpha^*p)dF(\theta|a)
\]
as well, given that $a$ is fixed.

**Proof of Proposition 6:** The first-order condition of the principal’s action choice is given by (5.5) under a truthful mechanism, and by
\[
\int \pi(\theta) - t(b_m(\theta))(1 - \alpha^*p)dF_a(\theta|a) + \lambda \left[ \int v(t(\hat{m}(\theta)))(1 - \alpha^*p)dF_{aa}(x|a) - c''(a') \right] = 0
\]
in the alternative problem

**Problem 3**
\[
\max_{a,t} \int \pi(x) - t[1 - \alpha^*p]dF(x|a)
\]
\[
s.t. \quad \int v(t)(1 - \alpha^*p)dF(x|a) \geq c(a) \quad (8.4)
\]
\[
\int v(t)(1 - \alpha^*p)dF_a(x|a) = c'(a) \quad (8.5)
\]
if not attempting to elicit the truth. By Lemma 3, $0 < \int [h(U) + \tau]dF(\theta|a) < \int t(\hat{m}(\theta))(1 - \alpha^*p)dF_a(\theta|a)$. Therefore we must have
\[
\left[ \int v(h(U) + \tau)dF_{aa}(x|a) - c''(a^{SB}) \right] < \left[ \int v(t(\hat{m}(\theta)))(1 - \alpha^*p)dF_{aa}(x|a) - c''(a') \right] < 0
\]
to satisfy the equality. So $a^{SB} \geq a'$ by convexity of $c(a)$.

**References**


