Optimal Monitoring Design*

George Georgiadis and Balazs Szentes†

January 31, 2018

Abstract

This paper considers a Principal-Agent model with hidden action in which the Principal can monitor the Agent by acquiring independent signals conditional on effort at a constant marginal cost. The Principal aims to implement a target effort level at minimal cost. The main result of the paper is that the optimal information acquisition strategy is a two-threshold policy and, consequently, the equilibrium contract specifies two possible wages for the Agent. This result provides a rationale for the frequently observed *single-bonus wage-contracts*.

1 Introduction

A general lesson from Contract Theory is that in order to induce a worker to exert effort, she should be rewarded for those output realizations which indicate high effort. Designing such an incentive scheme in practice can be more challenging for various reasons. For example, if a firm has many employees, profit reflects aggregate performance and it is hard to disentangle an individual worker's contribution from her coworkers'. Moreover, some aspects of performance (*e.g.*, quality of customer service), are difficult to quantify and measure. In such cases, firms can still monitor the workers by identifying variables which are informative about their effort. Then, by constructing performance measures based on these variables and offering wage plans contingent on these measures, firms can reduce agency costs. Indeed, firms devote significant resources to searching for effective ways to evaluate their employees; see for

*We are grateful to Dan Barron, Luis Rayo, Andy Skrzypacz, Philip Strack, and Bruno Strulovici, as well as to audiences at Indiana University, Northwestern University, the University of Michigan, and the University of Utah for helpful comments.

†G. Georgiadis: Kellogg School of Management, Northwestern University, Evanston, IL 60208, U.S.A., % g-georgiadis@kellogg.northwestern.edu;
B. Szentes: Department of Economics, London School of Economics, London, WC2A 2AE, U.K., b.szentes@lse.ac.uk

example, Mauboussin (2012), WorldatWork and Deloitte Consulting (2014) and Buckingham and Goodall (2015). The goal of this paper is a theoretical investigation of the optimal monitoring structure in the absence of freely available information about the Agent's effort.

In our specific setup, there is a single Agent who exerts one of continuum many efforts. The Principal can acquire arbitrarily many signals which are independent from one another conditional on the Agent's effort at a constant marginal cost. This is modeled by assuming that the Principal can observe a diffusion process with the drift being the Agent's effort at a cost proportional to the time at which the the Principal stops observing this process. A contract specifies a stopping time and a wage scheme which is contingent on the Principal's observation. The Agent is risk-averse and has limited liability. Our goal is to characterize the contract which induces the Agent to choose a target effort level at minimal cost while respecting limited liability.

Our main result shows that, under certain conditions on the Agent's utility function, the optimal contract features a binary wage scheme; *i.e.*, the Agent is paid a base wage, plus a fixed bonus if his performance is deemed sufficiently good. This provides a new rationale for single-bonus contracts which abound in practice (Holmström, 2016). This result also addresses the criticism that canonical Principal-Agent models generate optimal contracts that are sensitive to minutiae details of the Principal's exogenously given information (often assumed to be the output). Indeed, in these models, the Agent's wages depend on the likelihood ratios, so the optimal scheme is finely tuned to the Principal's signal structure. Only very particular distributions yield wage contracts that have any resemblance to the contracts observed in practice (Hart and Holmström, 1986). For example, in the canonical model, single-bonus contracts are optimal only if there are exactly two possible values of the likelihood ratio, which is an unlikely feature of a typical output distribution. In contrast, we show that the optimal signal structure has this property if it is endogenously determined by the Principal's information acquisition strategy.

Our analysis has two main building blocks. First, through a sequence of steps, we reformulate the problem of identifying an optimal contract to a *flexible information design problem*. In this new problem, the Principal's choice set is a set of distributions instead of the set of stopping rules. Second, we show that finding a solution to the information design problem is equivalent to characterizing an equilibrium in a *zero-sum game* played by the Principal and *Nature*. At the end, our main result is stated as an equilibrium characterization of this game. Below, we explain both of these ideas in detail.

Information Design. The Principal's problem can be decomposed into two parts: a stopping rule defining the information acquisition strategy and a wage function mapping from the Principal's observations to the Agent's monetary compensations. This wage can depend on the whole path of the diffusion. However, we argue that the optimal wage depends only on a one-dimensional real variable, henceforth referred to as the *score*. More precisely, we show that, for *any* given stopping rule, the cost-minimizing wage only depends on the value and the time of the Principal's last observation.[1] Since the optimal contract is incentive compatible and the drift of the diffusion is the target effort level, we show that the optimal wage can be expressed as the function of the (driftless) Brownian motion part of the stochastic process. Each stopping rule generates a distribution over the scores with zero mean. Using results from Skorokhod Embedding Theory, we show that the converse is also true: for any zero-mean distribution over the scores, there is a stopping time which generates this distribution. Recall that the Principal's information acquisition cost is the expectation of the stopping time. It turns out that the expectation of the stopping time generating a certain distribution is the variance of the distribution. Therefore, the Principal's contracting problem can be rewritten as an information design problem where she chooses a distribution over scores at cost equal to its variance (instead of a stopping time) and a wage function defined on scores.

The Zero-Sum Game. For any given distribution over scores F, the standard approach to solve for the optimal wage is to pointwise minimize the corresponding Lagrangian function; see for example, Bolton and Dewatripont (2005). Let λ denote the Lagrange multiplier corresponding to the incentive constraint and $L(\lambda, F)$ denote the value of the Lagrangian function evaluated at the cost-minimizing wage function. We show that strong duality holds, that is, the Principal's value for a given F is $\sup_\lambda L(\lambda, F)$. Since the Principal chooses F to minimize her overall cost, her problem can be written as $\inf_F \sup_\lambda L(\lambda, F)$. Instead of solving this inf sup problem, we characterize the solution of the corresponding sup inf problem. The key to this characterization is to observe that for each dual multiplier λ, the problem $\inf_F L(\lambda, F)$ is an unconstrained information design problem. Using the concavification arguments developed in Aumann and Perles (1965) and Kamenica and Gentzkow (2011), we show that there exists a solution among the binary distributions; *i.e.*, a distribution supported only on two points.

It remains to argue that the inf sup and sup inf problems are equivalent. This follows from von Neumann's Minimax Theorem (see von Neumann, 1928), if the following zero-sum game has a Nash equilibrium. The game is played by *Nature*, who chooses the dual multiplier to maximize the Lagrange function, L, and the Principal, who chooses a probability distribution over scores to minimize L. We prove that, under some conditions on the Agent's utility function, this game indeed has a unique equilibrium.[2] The aforementioned concavification

[1] The score is the counterpart of the derivative of the log-density function with respect to the Agent's effort in the canonical Principal-Agent model of Holmström (1979), which is a sufficient statistic for the optimal wage scheme.

[2] These conditions are satisfied if, for example, the Agent's utility exhibits constant absolute risk aversion,

argument implies that the Principal always has a best-response which is binary. We argue that the Principal's equilibrium distribution also has this feature, that is, there are only two values of the score which arise with positive probability. Consequently, the Agent's wage is also binary and hence, a single-bonus wage scheme is optimal.

We believe that considering such a zero-sum game might turn out to be useful to analyze a class of problems where the Principal not only designs the information structure but also determines other policy variables subject to certain constraints. In our model, this policy variable is the wage and the constraint requires incentive compatibility. In such environments, it is unclear how one can use the concavification arguments developed to solve unconstrained problems. However, analyzing the Principal's best responses in the zero-sum game is just an unconstrained information design problem. These best responses might have some robust features, as demonstrated in this paper, which then will be the feature of the equilibrium as well as the solution for the original optimization problem.

We establish an additional result which holds under different conditions on the Agent's utility function. We show that there exists a sequence of binary distributions and single-bonus wage schemes, which approximates the first-best outcome arbitrarily closely. In other words, the Principal's payoff in the limit is the same as it would be if effort was contractible. A contract in the sequence pays the Agent a base wage, plus a large bonus with a small probability. Intuitively, the condition on the Agent's utility function is satisfied if the Agent is not too risk-averse, and so it is not too expensive to motivate him with a large wage that he receives with a small probability. As an example, it is satisfied if the Agent's utility exhibits constant relative risk aversion with coefficient less than half.

How general is our main result regarding single-bonus contracts? As mentioned above, our assumption that the marginal cost of information is constant enables us to transform the Principal's problem to an information design problem where the cost of a distribution is its variance. Alternatively, we could have started with the information design problem where the Principal chooses a distribution over scores. We argue that as long as the cost of a distribution is a general convex moment of the distribution,[3] our main theorem holds, that is, the optimal wage scheme is binary. From this viewpoint, modeling the Principal's information acquisition with a diffusion can be considered as a micro-foundation for specifying the cost of a distribution as a moment.

We emphasize that our main result regarding single-bonus contracts is, at least partially, due to the Principal's ability to design the monitoring structure in a flexible way. In our setup, this is achieved by allowing the stopping time to depend on the already observed path.

or constant relative risk aversion with coefficient greater than half.

[3]That is, the cost is the expectation of a convex moment function.

If instead, one considers a less flexible, parametric class of monitoring structures, the optimal contract is unlikely to feature a binary wage scheme. For example, assume that the Principal can observe a normal signal around the Agent's effort at a cost of the variance. This would be the case in our model if the stopping rule was restricted to be deterministic. Then no matter what variance the Principal chooses, the range of her signal will be a continuum and each signal determines a different wage.

Related Literature

First and foremost, this paper is related to the literature on Principal-Agent problems under moral hazard. In the seminal work of Mirrlees (1976) and Holmström (1979), a Principal contracts with a risk-averse Agent. The Principal has access to a contractible signal which is informative about the Agent's effort. The authors characterize the wage contract which maximizes the Principal's profit subject to the Agent's incentive compatibility and participation constraints. Extensions of this model include settings in which the performance measure is not contractible, the Agent's effort is multi-dimensional and some tasks are easier to measure than others, or the Principal and the Agent interact repeatedly – see Bolton and Dewatripont (2005) for a comprehensive treatment. This literature almost always treats the Principal's signal as exogenously given. In reality, good performance measures are not readily available – they must be designed and optimized, and doing so is costly.

Dye (1986) analyzes a Principal-Agent model in which, after observing a (costless) signal that is informative of the Agent's effort, the Principal can acquire an additional costly signal. Just like in our model, the Principal's information acquisition strategy is contractible and the Agent's wage is contingent on all the observed signals. It is shown that, under certain conditions, the Principal acquires the additional signal only if the value of first signal is sufficiently low.[4] Feltham and Xie (1994) and Datar, Kulp and Lambert (2001) examine how a set of available performance measures should be weighed in an optimal linear wage scheme. It is shown, for example, that it may be optimal to ignore informative signals of effort.[5] Hoffman, Inderst and Opp (2017) and Li and Yang (2017) also analyze contracting problems with endogenous monitoring. The former considers a model in which the Principal observes signals that are informative of the Agent's one-shot effort over time, and designs the optimal deferred compensation scheme. Deferring compensation enables the Principal to obtain more accurate information and thus reduce agency costs, but because the Agent is less

[4] See also Townsend (1979), Baiman and Demski (1980), Young (1986) and Kim and Suh (1992) for related studies.

[5] Note that Holmström's *informativeness principle*, which asserts that any signal that is informative of the Agent's action should be incorporated into the optimal contract, does not apply if the Principal is restricted to linear contracts.

patient than the Principal, doing so also entails a cost. It is shown that with a risk-neutral and cash constrained Agent, the optimal compensation scheme pays out at either one, or two dates. Li and Yang (2017) studies a game in which the Agent's hidden action generates a signal, and the Principal chooses a partition of the state-space (at a cost that increases in the fine-ness of the partition), and a wage scheme that specifies the Agent's wage conditional on the cell of the partition in which the signal lies. Their main result shows that the optimal partition comprises of convex cells in the space of likelihood ratios.

As mentioned above, a criticism of standard Principal-Agent models is that they are of limited use for predicting the structure of incentive contracts observed in reality (Hart and Holmström, 1986 and Holmström, 2016). Motivated by the fact that in practice, contracts tend to be *simple* (*e.g.,* linear, single-bonus contracts, options, etc), many studies attempt to rationalize these simple contracts. For example, Holmström and Milgrom (1987) shows that in a dynamic setting in which the Agent chooses his effort repeatedly and his utility function exhibits constant absolute risk aversion, linear contracts are optimal. Carroll (2015) shows that linear contracts are also optimal with a risk-neutral Agent if the Principal is uncertain about the actions available to the Agent and has robust (*i.e.*, min-max) preferences. Barron, Georgiadis and Swinkels (2017) considers the case in which the Agent can game his contract by manipulating the distribution of output (*e.g.,* by taking on risk), and they show that with a risk-neutral and cash constrained Agent, a linear contract is optimal. While this literature has focused largely on the optimality of linear contracts, single-bonus contracts are also very common. Single-bonus contracts are shown to be optimal in Levin (2003) and Palomino and Prat (2003). The former studies a relational contracting model, while the latter considers a delegated portfolio management problem in which the Agent chooses both expected output and the riskiness of the portfolio (in the sense of second-order stochastic dominance) from a parametric family of distributions. In both settings, the optimality of single-bonus contracts relies on the Agent being risk-neutral.

Finally, our work is related to the information design literature, see for example Rayo and Segal (2010) and Kamenica and Gentzkow (2011). The latter considers a game between a *sender* who wishes to persuade a *receiver* to take a state-dependent action, and can do so by designing the receiver's signal. Using techniques developed in Aumann and Perles (1965), they provide a tractable framework for characterizing the optimal information structure. Together with Barron, Georgiadis and Swinkels (2017) and Boleslavsky and Kim (2017), this is one of the first papers to analyze an information design problem under moral hazard. Similar to us, Boleslavsky and Kim (2017) considers a Principal-Agent model under moral hazard in which the Principal can design the information structure, but state contingent payoffs (*i.e.*, wages) are exogenous. They show that the optimal distribution over posteriors contains either two or three posteriors in its support.

2 Model

There is a Principal and an Agent. The Agent exerts effort $a \in \mathbb{R}_+$ at cost $c(a)$. The Agent's choice of effort is unobservable, but it generates a diffusion X_t with drift a, that is, $dX_t = adt + dB_t$, where B_t is a standard Brownian motion with the corresponding canonical probability space (Ω, B, P) and $B_0 = 0$.[6] The Principal acquires information about the Agent's effort by observing this process. Information acquisition is costly and the Principal's cost is t if she chooses to observe this process until time t. The Agent's payoff is $u(w) - c(a)$ and the Principal's cost is $w + t$ if the Principal pays the Agent wage w. The function u is strictly increasing, strictly concave, and $\lim_{w\to\infty} u'(w) = 0$. The function c is strictly increasing and strictly convex.

The Principal can commit to a path-contingent stopping rule and a path-contingent wage scheme. To be more precise, a *contract* is a pair (τ, W), where τ is a stopping time and W is a mapping from *paths* to wages. Formally, $\tau : \Omega \to \mathbb{R}_+$ is a stopping time of the filtration generated by B_t, and $W : \Omega \times \mathbb{R}_+ \to \mathbb{R}$ is a measurable function. If the Principal stops information acquisition at time t and observes the path $\omega_t = \{\omega_i\}_{i\le t}$, then the Agent receives wage $W(\omega_t)$.[7] We assume that the Agent has limited liability, that is, the Principal faces a minimum wage constraint, $W \ge \underline{w}$, where $\underline{w} > -\infty$ and $u'(\underline{w}) < \infty$.[8]

The game played by the Principal and the Agent proceeds as follows. First, the Principal offers a contract. After observing this contract, the Agent chooses an effort level.[9] Then the Principal acquires information and pays the wage according to the offered contract. The Principal's objective is to induce the Agent to exert a target level of effort, a. Our goal is to characterize the optimal contract which achieves this goal at the lowest expected cost. Formally, we analyze the following constrained optimization problem:

$$\inf_{\tau, W} \mathbb{E}_{a^*}\left[W(\omega_\tau) + \tau\right] \tag{1}$$

$$\text{s.t. } a^* \in \arg\max_a \mathbb{E}_a\left[u(W(\omega_\tau))\right] - c(a) \tag{2}$$

$$W(\omega_\tau) \ge \underline{w}. \tag{3}$$

[6]In particular, $\Omega = C([0,\infty))$.

[7]Note that ω_t is a realization of the path of X until t.

[8]This assumption rules out the possibility that the Principal can approximate the first-best outcome by observing the process X_t for an arbitrarily short duration and offering a Mirrlees "shoot the Agent"contract (Bolton and Dewatripont, 2005).

[9]All results continue to hold if the contract must also satisfy a participation constraint. We omit it for simplicity.

3 Reformulating the Principal's Problem

This section accomplishes the following three goals:

First, we consider a relaxed version of the Principal's problem in which we replace the incentive constraint with the first-order condition corresponding to Agent's optimal choice of effort. Then we show that the only determinant of the wage, the so-called *score*, is the value of Brownian motion at the stopping time of the Principal. In other words, if the Principal stops acquiring information at time τ and observes the function ω_τ, she pays a wage which only depends on $\omega_\tau - a^*\tau$. Note that, since the Agent exerts effort a^* in equilibrium, this quantity is just the realization of $B_\tau\,(= X_\tau - a^*\tau)$.

Each stopping rule of the Principal results in a different distribution of the score with mean zero. Our second objective is to argue that the reverse is also true: the Principal can induce any zero-mean distribution over the scores by appropriately choosing her stopping rule. In addition, the Principal's expected cost is the variance of the distribution. This observation allows us to rewrite the Principal's problem as a flexible information design problem, where the Principal can choose any distribution over scores (instead of determining the stopping rule).

Finally, we consider the optimal wage scheme for any distribution over scores. In particular, we consider the Lagrangian corresponding to the Principal's information design problem, characterize the optimal wage as a function of the dual multiplier, and show that strong duality holds.

3.1 The Score

Our next objective is to show that, despite the Principal observing a path of the diffusion, the wage of the Agent depends only on the last value of the path. To this end, first observe that the Principal's problem can be decomposed into two parts: finding an optimal stopping rule and determining the wage scheme given the stopping time. In other words, the optimal wage structure minimizes the Principal's cost (subject to incentive compatibility and limited liability) for the optimal stopping rule. In this section, we take the Principal's information acquisition as given, and describe some properties of the optimal wages.

In order to better explain the derivation of the score in our setting, let us briefly recall how this object is derived in standard Principal-Agent models with continuous effort. Consider such a model where the Agent's effort a generates an observable output distribution defined by the cumulative distribution function (hereafter CDF) G_a and corresponding probability distribution function (hereafter PDF) g_a. A standard approach in the literature (see for example Bolton and Dewatripont, 2005) is to solve a relaxed problem where the incentive compatibility constraint is replaced by the weaker condition

$$\int u\left(w\left(z\right)\right)\left.\frac{\partial g_a\left(z\right)}{\partial a}\right\rfloor_{a=a^*}dz\geq c'\left(a^*\right).$$

This condition guarantees that the Agent prefers to exert a^* to a *local* downward deviation. The Principal's Lagrangian with this weaker constraint becomes

$$\int\left[w\left(z\right)-\lambda u\left(w\left(z\right)\right)\frac{\partial g_a\left(z\right)/\partial a\rfloor_{a=a^*}}{g_{a^*}\left(z\right)}\right]dG_{a^*}\left(z\right)+\lambda c'\left(a^*\right).$$

Note that, in addition to the endogenously determined $w\left(z\right)$, the integrand only depends on $\left(\partial g_a\left(z\right)/\partial a\rfloor_{a=a^*}\right)/g_{a^*}\left(z\right)$. Therefore, pointwise maximization of the Lagrangian yields that the optimal wage only depends on the derivative of the log-density. This quantity is often referred to as the (Fisher) *score*.

To provide an intuition for the score being the value of the Brownian motion at the stopping time in our setting, assume for simplicity that the Principal observes the diffusion up-to time t, that is, $\tau\equiv t$. If the Agent exerts effort a, then the density corresponding to the last observation, z, is $g\left(z|a\right)=e^{-(z-at)^2/2t}/\sqrt{2\pi t}$. Note that $\partial g\left(z|a\right)/\partial a=\left(z-at\right)g\left(z|a\right)$. So, if the Principal is restricted to make the wage dependent only on the last observation, the score becomes $z-a^*t$. Can the Principal benefit from additional observations? Suppose that the Principal makes the wage dependent on the last observation as well as the observation at $t/2$. Then $g\left(x,z|a\right)=\left[e^{-(x-at/2)^2/t}/\sqrt{\pi t}\right]\left[e^{-(z-x-at/2)^2/t}/\sqrt{\pi t}\right]$ and $\partial g\left(x,z|a\right)/\partial a=\left(z-at\right)g\left(x,z|a\right)$. Therefore, the score is still $z-a^*t$. In other words, even if the Principal could choose to make the wage depend on her observation at $t/2$ in addition to her last observation, she chooses not to do so.

In our model, given a stopping time, the Principal's observation about the Agent's effort is not a finite-dimensional object and cannot be described by a PDF. Nevertheless, Girsanov's Theorem characterizes a Radon-Nikodym derivative of the measure generated by a^* over the Principal's observations with respect to the measure generated by any other a. This enables us to express the Agent's deviation payoff for each effort a in terms of the measure generated by a^*. More precisely, Girsanov's Theorem implies that if the Agent exerts effort a, then her payoff is

$$\mathbb{E}_{a^*}\left[u\left(W\left(\omega_\tau\right)\right)e^{(a-a^*)B_\tau-\frac{1}{2}(a-a^*)^2\tau}\right]-c(a^*),\tag{4}$$

where the expectation is taken according to the measure over Ω generated by a^*. Differentiating this expression with respect to a, and evaluating it at $a=a^*$, we obtain the following relaxed incentive compatibility constraint

$$\mathbb{E}_{a^*}\left[u\left(W\left(\omega_\tau\right)\right) B_\tau\left(\omega_\tau\right)\right] \geq c'(a^*). \tag{5}$$

Using arguments similar to the ones explained in the previous paragraph, one can show that for any stopping rule τ, it is without loss of generality to condition wages only on B_τ, or equivalently, on the score $s_\tau := X_\tau - a^*\tau$.

Lemma 1. *Fix an arbitrary stopping rule τ, and consider the relaxed constrained optimization problem given by (1), (3) and (5). In an optimal contract, the Agent's wage only depends on s_τ.*

Proof. See the Appendix. □

In what follows, we characterize the solution to the relaxed problem, where the incentive compatibility constraint is replaced by (5). In Section 7, we provide conditions under which the solution to the original problem coincides with this relaxed one.

3.2 Flexible Information Design

Each stopping rule generates a probability distribution over scores. Let F_τ denote the CDF generated by the stopping time τ. Since B_t is a martingale, the expected value of the score is zero, that is, $F_\tau \in \mathcal{F} = \{F \in \Delta(\mathbb{R}) : E_F[s] = 0\}$. In what follows, we aim to rewrite the Principal's problem so that her choice set is $\mathcal{F}$ instead of the set of stopping rules. In other words, in this new problem, the Principal directly chooses a distribution over scores. A question that arises then is which distributions can be generated by some stopping rule and what is the corresponding cost. This is known as the Skorokhod embedding problem. The following lemma asserts that the Principal can generate essentially *any* distribution over scores by choosing an appropriate stopping time. Furthermore, the Principal's expected cost is the variance of the distribution. The following lemma is due to Root (1969) (Theorem 2.1) and Rost (1976) (Theorem 2).

Lemma 2. *Pick any $F \in \mathcal{F}$ such that $\mathbb{E}_F[s^2] < \infty$.*
(i) There exists a stopping time τ such that $F_\tau = F$ and $\mathbb{E}[\tau] = \mathbb{E}_F[s^2]$, and
(ii) if $F_{\tau'} = F$ for stopping time τ' then $\mathbb{E}[\tau] \leq \mathbb{E}[\tau']$.

The previous two lemmas allow us to reformulate the Principal's problem as an information design problem. Formally,

$$\inf_{F\in\mathcal{F},\widetilde{W}} \mathbb{E}_F\left[\widetilde{W}(s)+s^2\right] \tag{Obj}$$

$$\text{s.t. } \mathbb{E}_F\left[su\left(\widetilde{W}(s)\right)\right] \geq c'(a^*), \tag{IC}$$

$$\widetilde{W}(s) \geq \underline{w} \text{ for all } s\in\mathbb{R}. \tag{LL}$$

To be precise, the previous lemma requires F to have finite variance. We ignore this constraint and show that the solution of this problem satisfies this constraint.

If $\widetilde{W}$ was set to be the optimal wage scheme, then finding the optimal F in the previous problem becomes a pure information design problem. Of course, the optimal distribution must still satisfy the two constraints, (IC) and (LL). We intend to use standard techniques in information design developed to analyze *unconstrained* optimization problems. Therefore, our next goal is to eliminate the constraints by considering the corresponding Lagrangian function.

3.3 Optimal Wages and Strong Duality

As mentioned above, the Principal's problem can be decomposed into two parts: finding an optimal distribution over scores and determining the wage scheme given this distribution. This section focuses on the second part: for each F we characterize the wage scheme which minimizes the Principal's cost subject to (local) incentive compatibility. Formally, for each $F\in\mathcal{F}$, we consider

$$\begin{aligned}\inf_{\widetilde{W}} \quad & \mathbb{E}_F\left[\widetilde{W}(s)+s^2\right] \\ \text{s.t.} \quad & \text{(IC) and (LL).}\end{aligned} \tag{6}$$

Note that this is a standard Principal-Agent problem as in Holmström (1979), except that above, we have a limited liability constraint instead of an individual rationality constraint, and F is a probability distribution over scores instead of outputs. Let $\Pi(F)$ denote the value of this problem.

The Lagrangian function corresponding to this problem can be written as

$$L(\lambda,F)=\inf_{\widetilde{W}(\cdot)\geq\underline{w}}\int\left[\widetilde{W}(s)-\lambda su(\widetilde{W}(s))+s^2\right]dF(s)+\lambda c'(a), \tag{7}$$

where $\lambda\geq 0$ is the dual multiplier associated with (IC). To solve this problem, note that the first-order condition corresponding to the pointwise minimization of the integrand is

$\lambda s u'(w) = 1$. If the solution of this equation, w, is larger then $\underline{w}$ then this w is going to be the optimal wage at s. Otherwise, the optimal wage is $\underline{w}$. To summarize, the wage scheme that minimizes the value of the integral on the right-hand side of (7) is defined by the following equation:

$$w(\lambda, s) = \begin{cases} \underline{w} & \text{if } s \leq s_*(\lambda) \\ u'^{-1}\left(\frac{1}{\lambda s}\right) & \text{if } s > s_*(\lambda)\,, \end{cases} \tag{8}$$

where $s_*(\lambda)$ is the critical score at which the solution of the first-order condition is exactly $\underline{w}$, that is, $s_*(\lambda) = 1/\left[\lambda u'(\underline{w})\right]$. The following lemma shows that if $w(\lambda, \cdot)$ satisfies (IC) for some $\lambda \geq 0$, then this wage scheme is uniquely optimal. Moreover, strong duality always holds.

Lemma 3. *(i) Suppose that there exists a* $\lambda^* \in \mathbb{R}_+$ *such that*

$$\int s u(w(\lambda^*, s)) dF(s) = c'(a^*)\,,$$

where $w(\lambda^*, s)$ *is given in (8). This wage scheme uniquely solves (6) subject to (IC) and (LL). (ii) Strong duality holds; i.e.,* $\sup_{\lambda \in \mathbb{R}_+} L(\lambda, F) = \Pi(F)$.

Proof. See the Appendix. □

The first statement is similar to Proposition 1 of Jewitt, Kadan and Swinkels (2008). The difference is that wages are specified as a function of output in their setting, whereas in our model, they depend on the realized score.

Recall that throughout this section, we fixed the distribution over the scores, F, and characterized the corresponding optimal wage scheme. Of course, the Principal also chooses this distribution to minimize her cost, that is, she solves $\inf_{F \in \mathcal{F}} \Pi(F)$. Part (ii) of this lemma enables us to rewrite this problem as

$$\inf_{F \in \mathcal{F}} \sup_{\lambda \in \mathbb{R}_+} L(\lambda, F)\,. \tag{9}$$

It turns out to be difficult to characterize the F that solves this problem. The reason is that there is a different λ corresponding to each possible F, and hence, it is hard to identify the change in $\sup_{\lambda \in \mathbb{R}_+} L(\lambda, F)$ due to a change in F. In the next section, we show that solving the corresponding $\sup \inf$ problem is simpler and we investigate the circumstances under which the two problems are equivalent.

4 The Zero-Sum Game

Our next objective is to define a zero-sum game and show that, if there exists an equilibrium in this game, the inf sup problem in (9) is equivalent to

$$\sup_{\lambda \in \mathbb{R}_+} \inf_{F \in \mathcal{F}} L(\lambda, F). \tag{10}$$

We are able to characterize the solution of this sup inf problem. Indeed, for any λ, $\inf_{F \in \mathcal{F}} L(\lambda, F)$ is an information design problem akin to that in Kamenica and Gentzkow (2011). We show that for any λ, if $\inf_{F \in \mathcal{F}} L(\lambda, F)$ has a solution, then there is an optimal F which is either a *two-point distribution* (*i.e.*, its support has two elements), or F is degenerate, specifying an atom of size one at zero.

In what follows, we first formally define the zero-sum game. Then, using arguments from the theory of zero-sum games, we prove that equilibrium existence implies the equivalence of (9) and (10). Finally, we explain how *two-point distributions* arise as best-responses in this game. This last observation is crucial to our main result according to which the optimal wage scheme is binary.

The Game.— There are two players, the Principal and Nature. The action space of the Principal is $\mathcal{F}$ and the action space of Nature is $\mathbb{R}_+$. Furthermore, Nature's payoff is $L(\lambda, F)$. That is, the Principal chooses a probability distribution $F \in \mathcal{F}$ to minimize $L(\lambda, F)$, whereas nature chooses the dual multiplier λ to maximize $L(\lambda, F)$.

The following lemma, which is due to von Neumann (1928), shows that a Nash equilibrium of this game corresponds to a solution to both (9) and (10).

Lemma 4. *Suppose that $\{\lambda^*, F^*\}$ is a Nash equilibrium in the zero-sum game defined above. Then*

$$\sup_{\lambda \geq 0} \inf_{F \in \mathcal{F}} L(\lambda, F) = \inf_{F \in \mathcal{F}} \sup_{\lambda \geq 0} L(\lambda, F) ,$$

and $w(\lambda^*, \cdot)$ and F^* solves (Obj)-(LL).

Proof.
If $\{\lambda^*, F^*\}$ is an equilibrium in the zero-sum game defined above, then

$$\inf_{F \in \mathcal{F}} \sup_{\lambda \geq 0} L(\lambda, F) \leq \sup_{\lambda \geq 0} L(\lambda, F^*) = L(\lambda^*, F^*) = \inf_{F \in \mathcal{F}} L(\lambda^*, F) \leq \sup_{\lambda \geq 0} \inf_{F \in \mathcal{F}} L(\lambda, F),$$

where the two equalities hold because λ^* and F^* are best-responses to each other. Since $\inf_{F \in \mathcal{F}} \sup_{\lambda \geq 0} L(\lambda, F) \geq \sup_{\lambda \geq 0} \inf_{F \in \mathcal{F}} L(\lambda, F)$ always holds, the previous inequality chain implies that

$$\sup_{\lambda\geq 0} L\left(\lambda, F^{*}\right) = \inf_{F\in\mathcal{F}} \sup_{\lambda\geq 0} L\left(\lambda, F\right) = L\left(\lambda^{*}, F^{*}\right) = \inf_{F\in\mathcal{F}} L\left(\lambda^{*}, F\right).$$

The first equality asserts that F^* is a solution to the original mini-max problem, (9). The last equality asserts that the solution to the original problem, F^*, is also a best-response to the multiplier λ^*.

Finally, it follows from Lemma 3 that $w(\lambda^*, \cdot)$ and F^* solves (Obj)-(LL).

If the first-order approach is valid, which we verify in Section 7, then the wage scheme $w(\lambda^*, \cdot)$ and the stopping rule that corresponds to the distribution F^* solves the Principal's original problem. □

Two-point Distribution.— Next, we argue that the Principal's best-response is either a two-point distribution or the degenerate distribution placing an atom of size one at zero. Furthermore, we argue that the latter case cannot arise in equilibrium. To this end, recall from (7) that the payoffs can be expressed as an expectation, that is,

$$L(\lambda, F) = \mathbb{E}_F\left[Z(\lambda, s)\right], \text{ where } Z(\lambda, s) = w(\lambda, s) - \lambda\left[su(w(\lambda, s)) - c'(a^*)\right] + s^2. \tag{11}$$

Then the problem of finding the Principal's best-response against λ can be written as

$$\inf_{F\in\mathcal{F}} \mathbb{E}_F\left[Z(\lambda, s)\right].$$

The solution to this problem can be characterized as follows by using standard arguments in information design, see Aumann and Perles (1965) and Kamenica and Gentzkow (2011). First, let $Z^c\left(\lambda, \cdot\right)$ denote the convexification of $Z\left(\lambda, \cdot\right)$ in s; *i.e.*,

$$Z^c(\lambda, s) = \inf_{\underline{s},\overline{s}\in\mathbb{R},\ \alpha\in[0,1] \text{ s.t. } \alpha\underline{s}+(1-\alpha)\overline{s}=s} \left\{\alpha Z(\lambda, \underline{s}) + (1-\alpha)\, Z(\lambda, \overline{s})\right\}, \tag{12}$$

Note that for any $F \in \mathcal{F}$,

$$\mathbb{E}_F\left[Z(\lambda, s)\right] \geq \mathbb{E}_F\left[Z^c(\lambda, s)\right] \geq Z^c(\lambda, 0),$$

where the first inequality follows because $Z(\lambda, s) \geq Z^c(\lambda, s)$, and the second one follows from Jensen's inequality. This inequality implies that $Z^c(\lambda, 0)$ is a lower bound on the Principal's payoff. Next, we explain that the Principal can achieve this bound by considering the following two cases.

If $Z(\lambda, 0) > Z^c(\lambda, 0)$, then the point $(0, Z^c\left(\lambda, 0\right))$ lies on the line-segment defining Z^c on

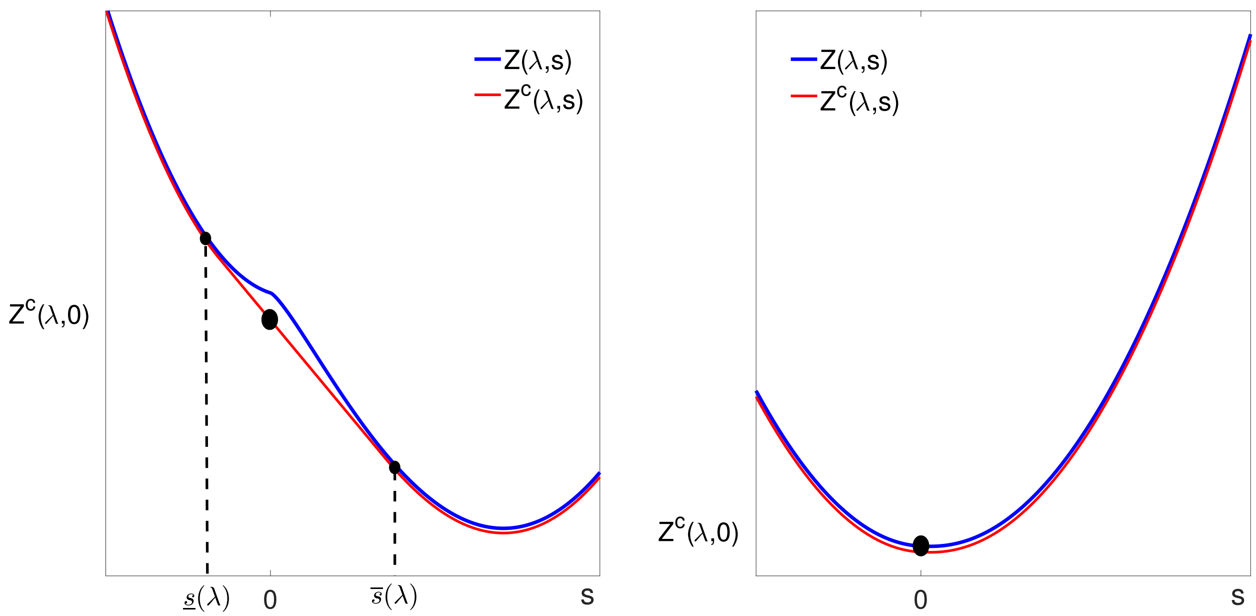

Figure 1 (a) Figure 1 (b)

Figure 1: Illustration of the Principal's best response.

the non-convex region around 0, as illustrated in Figure 1(a).[10] The point $(0, Z^c(\lambda, 0))$ is a convex combination of $(\underline{s}, Z(\lambda, \underline{s}))$ and $(\overline{s}, Z(\lambda, \overline{s}))$ for some $\underline{s} < 0 < \overline{s}$, that is, there exists $\alpha \in (0, 1)$ such that

$$\alpha (\underline{s}, Z(\lambda, \underline{s})) + (1 - \alpha)(\overline{s}, Z(\lambda, \overline{s})) = (0, Z^c(\lambda, 0)) . \tag{13}$$

Consider now the probability distribution, $\widehat{F}$, defined by the weights in this convex combination over $\{\underline{s}, \overline{s}\}$; *i.e.*, an atom of size α at $\underline{s}$ and an atom of size $(1 - \alpha)$ at $\overline{s}$. Equation (13) implies that $\alpha\underline{s} + (1 - \alpha)\overline{s} = 0$ which means that $\widehat{F}$ is feasible for the Principal, that is, $\widehat{F} \in \mathcal{F}$. Equation (13) also implies that

$$\alpha Z(\lambda, \underline{s}) + (1 - \alpha) Z(\lambda, \overline{s}) = \mathbb{E}_{\widehat{F}}[Z(\lambda, s)] = Z^c(\lambda, 0) ,$$

which means that the lower bound, $Z^c(\lambda, 0)$, is attained by the distribution $\widehat{F}$. Therefore, $\widehat{F}$ is a best-response of the Principal.

If $Z^c(\lambda, 0) = Z(\lambda, 0)$, then the lower bound can be trivially attained by the degenerate distribution which places probability only on zero, as illustrated in Figure 1(b). However, this latter case cannot arise in equilibrium. To see this, suppose that the distribution F is

[10] To be precise, the infimum need not be attained by any distribution. In this case, the Principal achieves $Z^c(\lambda, 0)$ only as a limit of two-point distributions.

degenerate and specifies an atom of size one at zero. Then, by (8), $w(\lambda, 0) = \underline{w}$ and hence, Nature's payoff becomes $\underline{w} + \lambda c'(a^*)$, by (11). Since $c'(a^*) > 0$, this quantity is strictly increasing in λ, and therefore, Nature does not have a best-response. This, in turn, implies that the degenerate distribution cannot arise in an equilibrium.

The observation that the Principal has a two-point distribution best response against Nature's equilibrium multiplier is the key observation for our main result. Indeed, any two-point distribution corresponds to an information acquisition strategy of the Principal which generates only two possible values of the score. As a consequence, the Agent receives only two possible wages. Note, however, that the argument in the previous paragraphs by no mean implies that the Principal does not have a best-response which is supported on more than two points. Of course, it is possible that the line connecting the points $(\underline{s}, Z(\lambda, \underline{s}))$ and $(\overline{s}, Z(\lambda, \overline{s}))$ contains other points $(s, Z(\lambda, s))$. In such a scenario, the point $(0, Z(\lambda, 0))$ is also a convex combination of (weakly) more than three such points on the line and each of these convex combinations define a best-response of the Principal.

To summarize, in order to show that the Principal's optimal contract specifies two possible wages we need to prove that (i) both the Principal's and Nature's best responses are unique, and (ii) an equilibrium in our zero-sum game exists.

5 The Main Theorem

Our main theorem provides sufficient conditions on the Agent's utility function for the existence of a unique equilibrium in the zero-sum game described in the previous section.

Theorem 1. *Suppose that the Agent's utility function u satisfies the following two conditions:*

(i) $[u']^3/u''$ is strictly increasing and

(ii) $\lim_{w\to\infty} [u'^3(w)/u''(w)] = 0$.

Then there exists a unique equilibrium $(\lambda^*, F^*) \in \mathbb{R}_+ \times \mathcal{F}$ in the zero-sum game, and F^* is a two-point distribution.

Let us explain the implication of this theorem to the original contracting problem. By Lemma 4, the equilibrium (λ^*, F^*) defines the solution for the constrained information design problem in (Obj). That is, the optimal distribution over scores is F^* and the wage scheme is $\{w(\lambda^*, s)\}_s$. By Lemma 2, this wage scheme and the optimal stopping rule generating F^* over the scores solves the Principal's relaxed problem (1) subject to (3) and (5). According to this theorem, F^* is a two-point distribution and let $\{\underline{s}, \overline{s}\}$ denote its support such that $\underline{s} < 0 < \overline{s}$. Then the Principal's optimal information acquisition strategy can be defined by the stopping rule in which she observes the diffusion process $X_t = at + B_t$ until the first time it hits $a^*t + \underline{s}$ or $a^*t + \overline{s}$. If the Agent exerts a^*, the value of the Brownian motion is either $\underline{s}$ or $\overline{s}$ at the stopping time, so the Principal observes two possible scores. Finally, the Principal pays the Agent $\underline{w}$ if she observed $\underline{s}$, and pays him $w(\lambda^*, \overline{s})$ if she observed $\overline{s}$. In other words, the Agent receives a base wage of $\underline{w}$ and bonus of $w(\lambda^*, \overline{s}) - \underline{w}$ if the information gathered is favorable, which is just a single-bonus contract.

Condition (i) is familiar from the literature on moral hazard problems with continuous effort. Indeed, one of the sufficient conditions guaranteeing that the first-order approach is valid (*i.e.*, that the global incentive constraint can be replaced by a local one), is that the function $\rho(z) = u(u'^{-1}(1/z))$ is concave (see Jewitt, 1988). It is not hard to show that condition (i) is equivalent to ρ being strictly concave.[11] Condition (ii) requires the derivative of ρ to vanish at infinity.

We prove this theorem in Section 5.1. In what follows, we discuss how restrictive the two conditions of the theorem are by examining whether they are satisfied by familiar parametric families of utility functions. Then we provide a sketch of the proof of the theorem.

Commonly Used Utility Function.—Consider first utility functions exhibiting constant absolute risk aversion (CARA), that is,

$$u(w) = -e^{-\alpha w}, \tag{14}$$

where $\alpha\,(>0)$ is the coefficient of absolute risk-aversion. In this case, $[u'(w)]^3/u''(w) = -\alpha e^{-\alpha w}$, so conditions (i) and (ii) are satisfied for all α.

Suppose now that the Agent's utility function exhibits constant relative risk aversion (CRRA), that is,

$$u(w) = \frac{w^{1-\gamma}}{1-\gamma}, \tag{15}$$

where $\gamma \in (0,1)$ is the coefficient of relative risk-aversion. In this case, $[u'(w)]^3/u''(w) = -\left[w^{1-2\gamma}\right]/\gamma$. Therefore, conditions (i) and (ii) of the theorem are satisfied if and only if $\gamma > 1/2$. What happens if $\gamma < 1/2$? Observe that the Principal's cost is bounded from below by $\underline{w}$. In Section 6, we show that the Principal can induce the Agent to exert (any) a^* at the minimum cost of $\underline{w}$. To be more precise, for each $\varepsilon > 0$, we construct an incentive compatible single-bonus contract such that the Principal's payoff from this contract is less than $\underline{w} + \varepsilon$.

If the Agent has a logarithmic utility function, $u(w) = \log w$, then $[u'(w)]^3/u''(w) = -1/w$, so the conditions of the theorem are satisfied.

More generally, if the Agent's utility function exhibits hyperbolic absolute risk aversion

[11] To see this, note that, denoting $u'^{-1}(1/z)$ by $f(z)$, $\rho'(z)$ can be expressed as $-\left[u'(f(z))\right]^3/u''(f(z))$. Since f is strictly increasing, ρ is strictly decreasing if and only if condition (i) holds.

(HARA), and so is of the form

$$u(w) = \frac{1-\gamma}{\gamma}\left(\frac{\alpha w}{1-\gamma}+\beta\right)^{\gamma}, \tag{HARA}$$

then conditions (i) and (ii) are satisfied if $\alpha > 0$, $\gamma < 1/2$, and $\beta > -\alpha \underline{w}/(1-\gamma)$.

Proof-Sketch.— Let us first explain the last statement of the theorem; *i.e.*, provided that an equilibrium exists, the Principal's equilibrium distribution is a two-point distribution. Applying the convexification argument described in the previous section, we show that the Principal's best-response is either a two-point distribution or the degenerate one. This is the part of the proof where we use assumptions (i) and (ii) of the hypothesis of the theorem. As we will show, a consequence of these assumptions is that the function Z looks like either the one depicted in Figure 1(a) or the one in Figure 1(b). More precisely, if λ is large then the function $Z(\lambda, \cdot)$ is convex-concave-convex whereas for small λ, this function is convex. This observation implies that the Principal's best-response is essentially unique and there are no best-response distributions of the Principal which are supported on at least three points. Since the equilibrium distribution cannot be degenerate, this implies that the Principal's equilibrium distribution is a two-point distribution.

Let us now explain the two main steps of the equilibrium-existence result in Theorem 1. The first step is to characterize some properties of Nature's best-responses. Using standard arguments from Lagrangian optimization, we show that, unless F is the degenerate distribution, the best-response λ against F is defined by the incentive constraint (IC), that is, Nature chooses the multiplier so that the incentive constraint binds. The second step is to show that if λ is small, then the incentive constraint (IC) evaluated at the Principal's best-response is violated. In contrast, if λ is large, then the incentive constraint (IC) evaluated at the Principal's best-response is slack. Then we use the Intermediate Value Theorem to conclude that there exists a unique λ^* at which the incentive constraint at the Principal's best response, say F^*, binds. As mentioned above, Nature's best response is characterized precisely by this binding incentive constraint. Therefore, λ^* is also best-response against F^*.

5.1 Proof of Theorem 1

Towards proving Theorem 1, we establish a series of lemmas, which enable us to construct a unique equilibrium in the zero-sum game described in Section 4.

5.1.1 Nature's Best Response

First, we show that Nature best-responds to a distribution F by choosing λ so that the Agent's incentive constraint (IC) binds, that is,

$$\int su\left(w\left(\lambda, s\right)\right) dF\left(s\right) = c'(a^*). \tag{16}$$

Formally, we state the following

Lemma 5. *(i) If (16) has a solution, then it is unique and defines Nature's best-response, λ_F.*

(ii) If (16) does not have a solution, then Nature does not have a best-response.

Regarding part (ii), we point out that Nature has no best response if her objective function is strictly increasing in λ, so she can improve on any λ by choosing a larger one. This is the case, for example, if the Principal's distribution is degenerate; *i.e.*, $F(s) = \mathbb{I}_{\{s\geq 0\}}$. It can be shown that if $\lim_{w\to\infty} u\left(w\right) = \infty$, then (16) always has a solution unless F is the degenerate distribution.

Proof.

Recall that if the Principal chooses F then Nature's problem is

$$\sup_{\lambda\geq 0} \mathbb{E}_F\left[Z(\lambda, s)\right] , \tag{17}$$

where $Z(\lambda, s)$ is defined in (11). Since the wage scheme $w\left(\lambda, \cdot\right)$ is optimally chosen for λ (see (8)), the Envelope Condition implies that

$$\frac{\partial \mathbb{E}_F\left[Z(\lambda, s)\right]}{\partial \lambda} = -\int su\left(w\left(\lambda, s\right)\right) dF\left(s\right) + c'(a^*)\,. \tag{18}$$

Note that the first-order condition corresponding to this derivative is just (16).

The second-order condition corresponding to (17) is

$$\frac{\partial^2 \mathbb{E}_F\left[Z(\lambda, s)\right]}{\partial \lambda^2} = \int_{s_*(\lambda)}^{\infty} s^2 \frac{\left[u'(w(\lambda, s))\right]^3}{u''(w(\lambda, s))} dF(s) \leq 0\,,$$

where the inequality follows from u being strictly increasing and strictly concave. Notice that the inequality is strict whenever $F\left(s_*(\lambda)\right) < 1$, that is, $s > s_*(\lambda)$ with positive probability.

The previous displayed inequality implies that Nature's objective function, $\mathbb{E}_F\left[Z(\lambda, s)\right]$, is concave in λ. Therefore, if (17) has an interior solution then it is defined by the first-order condition (16). This completes the proof of part (i).

To see part (ii), note that

$$\lim_{\lambda\to 0}\frac{\partial \mathbb{E}_F\left[Z(\lambda,s)\right]}{\partial\lambda} = -\lim_{\lambda\to 0}\int su\left(w\left(\lambda,s\right)\right)dF\left(s\right)+c'(a^*) = u\left(\underline{w}\right)\int sdF\left(s\right)+c'(a^*) = c'(a^*),$$

where the second equality follows from the fact that $w(\lambda, s)$ is decreasing in λ and and converging to $\underline{w}$ for all s and from Lebesgue's Monotone Convergence Theorem. The last equality follows from $F \in \mathcal{F}$, that is, F has zero expectation. The previous equation implies that that Nature's objective function, $\mathbb{E}_F[Z(\lambda, s)]$, is strictly increasing in λ at zero. This means that if (16) does not have a solution, Nature's objective function is strictly increasing for all $\lambda \geq 0$ and hence, Nature has no best response. □

5.1.2 Principal's Best Response

This section is devoted to the characterization of the Principal's best response. Recall that for a given λ, the Principal's problem is $\inf_{F\in\mathcal{F}}\mathbb{E}_F[Z(\lambda,s)]$. As mentioned above, this is an information design problem and we solve it by using standard convexification arguments. Of course, the solution depends on the shape of the function $Z(\lambda, \cdot)$. The next lemma shows that if the two conditions of Theorem 1 are satisfied, then this function is strictly convex for small values of λ, whereas it is convex-concave-convex for larger values of λ. Then we show that, in the former case, the Principal's best-response is degenerate, placing all the probability mass at zero and, in the latter case, the best-response is a binary distribution.

Throughout this section we maintain assumptions (i) and (ii) in the statement of Theorem 1.

Lemma 6. *There exists a $\lambda_c > 0$ such that*
(i) if $\lambda \leq \lambda_c$ then $Z(\lambda, s)$ is strictly convex in s and
(ii) if $\lambda > \lambda_c$ then there exists a $\widetilde{s} > s_*(\lambda)$ such that[12]

$$Z_{22}(\lambda,s)=\begin{cases}>0 & \text{if } s<s_*(\lambda)\ ,\\ <0 & \text{if } s_*(\lambda)<s<\widetilde{s}\ ,\text{ and}\\ >0 & \text{if } s>\widetilde{s}\,.\end{cases}\tag{19}$$

Proof.

Suppose first that $s < s_*(\lambda)$. Then $Z_2(\lambda, s) = -\lambda u(\underline{w}) + 2s$ and

$$Z_{22}\left(\lambda,s\right)=2>0,$$

[12] We use $Z_i(\lambda, s)$ and $Z_{ii}(\lambda, s)$ to denote the first and second derivative of $Z(\lambda, s)$ with respect to its i^{th} argument, respectively.

yielding the convexity of $Z(\lambda, s)$ for $s \leq s_*(\lambda)$ and, in particular, the first line of (19). If $s \geq s_*(\lambda)$ then $Z_2(\lambda, s) = -\lambda u(w(\lambda, s)) + 2s$ and

$$Z_{22}(\lambda, s) = \lambda^2 \frac{[u'(w(\lambda, s))]^3}{u''(w(\lambda, s))} + 2.$$

Note that $Z_{22}(\lambda, s)$ is strictly increasing in s because $w(\lambda, s)$ is is strictly increasing in s and $[u'(w)]^3 / u''(w)$ is strictly increasing in w by assumption. Note that at $s = s_*(\lambda)$, this second derivative is

$$Z_{22}(\lambda, s_*) = \lambda^2 \frac{[u'(\underline{w})]^3}{u''(\underline{w})} + 2.$$

Let us define λ_c by

$$\lambda_c = \sqrt{\frac{-2u''(\underline{w})}{[u'(\underline{w})]^3}},$$

and note that $Z_{22}(\lambda_c, s_*(\lambda_c)) = 0$.

If $\lambda \leq \lambda_c$ then $Z_{22}(\lambda, s_*(\lambda)) \geq 0$. Since $Z_{22}(\lambda, s)$ is increasing in s, the statement of the lemma follows.

If $\lambda > \lambda_c$ then $Z_{22}(\lambda, s_*(\lambda)) < 0$. As we mentioned before, the derivative $Z_{22}(\lambda, s)$ is strictly increasing. Furthermore, by assumption, this derivative becomes positive because $w(\lambda, s)$ goes to infinity as s goes to infinity. Since, $Z_{22}(\lambda, s)$ is continuous at $s > s_*(\lambda)$, there exists a unique $\widetilde{s}$ at which $Z_{22}(\lambda, \widetilde{s}) = 0$. □

We will argue that the value of the Principal's problem, $\inf_{F \in \mathcal{F}} \mathbb{E}_F[Z(\lambda, s)]$, is just the convexified $Z(\lambda, \cdot)$ evaluated at $s = 0$ (see Figure 1). Recall that this convexification, $Z^c(\lambda, \cdot)$, is defined by (12). Lemma 6 enables us to describe $Z^c(\lambda, \cdot)$. If $\lambda \leq \lambda_c$, then, by part (i) of Lemma 6, $Z(\lambda, \cdot)$ is convex, so $Z^c(\lambda, s) = Z(\lambda, s)$ for all s. If $\lambda > \lambda_c$, then, by part (ii) of Lemma 6, Z is convex-concave-convex. The convexification then only affects $Z(\lambda, \cdot)$ around the concave region. The function $Z^c(\lambda, \cdot)$ is linear on an interval around the concave region of $Z(\lambda, \cdot)$ and, otherwise, it coincides with $Z(\lambda, \cdot)$ (see the left panel on Figure 1). Formally, there exist $\underline{s}(\lambda)$ and $\overline{s}(\lambda) \geq \underline{s}(\lambda)$ such that

$$Z^c(\lambda, s) = \begin{cases} Z(\lambda, s) & \text{if } s \notin [\underline{s}(\lambda), \overline{s}(\lambda)] \\ Z(\lambda, \underline{s}(\lambda)) + (s - \underline{s}(\lambda)) Z_2(\lambda, \underline{s}(\lambda)) & \text{if } s \in [\underline{s}(\lambda), \overline{s}(\lambda)]. \end{cases} \tag{20}$$

Next, we show that the Principal's best-response is degenerate if either $\lambda \leq \lambda_c$ or if Z is not affected by the convexification around zero; *i.e.*, $0 \notin [\underline{s}(\lambda), \overline{s}(\lambda)]$. Otherwise, the Principal's best-response is a binary distribution.

Lemma 7. *For any* $\lambda\,(\geq 0)$, $\min_{F\in\mathcal{F}}\mathbb{E}_F\left[Z(\lambda,s)\right] = Z^c(\lambda,0)$. *In addition, the Principal's best-response,* F_λ, *is unique, and:*

(i) If $Z^c(\lambda,0) = Z(\lambda,0)$, *then* $F_\lambda(s) = \mathbb{I}_{\{s\geq 0\}}$.

(ii) If $Z^c(\lambda,0) < Z(\lambda,0)$, *then supp* $(F_\lambda) = \{\underline{s}(\lambda),\,\overline{s}(\lambda)\}$.

Observe that in part (ii), the Principal's best-response can be explicitly expressed in terms of its support by the following equation:

$$F_\lambda(s) = \begin{cases} 0 & \text{if } s < \underline{s}(\lambda) \\ \frac{\overline{s}(\lambda)}{\overline{s}(\lambda)-\underline{s}(\lambda)} & \text{if } s \in [\underline{s}(\lambda),\overline{s}(\lambda)) \\ 1 & \text{if } s \geq \overline{s}(\lambda). \end{cases} \tag{21}$$

This CDF has two jumps at $\underline{s}(\lambda)$ and at $\overline{s}(\lambda)$, so only these two points occur with positive probability. The size of each jump is determined by the requirement that F_λ has zero expectation. Observe that since $Z^c(\lambda,0) < Z(\lambda,0)$ in this case, it must be that $\underline{s}(\lambda) < 0 < \overline{s}(\lambda)$.

Proof.

Fix some $\lambda \geq 0$. By construction, $Z(\lambda,s) \geq Z^c(\lambda,s)$, so for all $F \in \mathcal{F}$, we have $\mathbb{E}_F[Z(\lambda,s)] \geq \mathbb{E}_F[Z^c(\lambda,s)] \geq Z^c(\lambda,\mathbb{E}_F[s]) = Z^c(\lambda,0)$, where the last inequality follows from the fact that $Z^c(\lambda,\cdot)$ is convex and Jensen's inequality. Therefore, $Z^c(\lambda,0)$ poses a lower bound on $\mathbb{E}_F[Z(\lambda,s)]$ for any $F \in \mathcal{F}$.

Part (i) follows trivially by noting that if $Z^c(\lambda,0) = Z(\lambda,0)$, then $\mathbb{E}_F\left[Z(\lambda,s)\right] = Z^c(\lambda,0)$ for $F(s) = \mathbb{I}_{\{s\geq 0\}}$. If $Z^c(\lambda,0) < Z(\lambda,0)$, then it follows from Lemma 6 and the definition of $Z^c(\lambda,\cdot)$ that there exist there exist s_L, s_H and $p \in (0,1)$ such that $(1-p)s_L + ps_H = 0$ and $Z^c(\lambda,0) = (1-p)Z(\lambda,s_L) + pZ(\lambda,s_H) = \mathbb{E}_{F_\lambda}\left[Z(\lambda,s)\right]$, where F_λ is of the form given in (21) with $\overline{s}(\lambda) = s_H$ and $\underline{s}(\lambda) = s_L$. Thus, we have shown that $\min_{F\in\mathcal{F}}\mathbb{E}_F[Z(\lambda,s)] = Z^c(\lambda,0)$ and established part (ii). □

The following lemma shows that $\underline{s}\,(\lambda)$ and $\overline{s}\,(\lambda)$ are continuous in λ, and both $\underline{s}\,(\lambda)$ and $\overline{s}\,(\lambda)$ converge to $s_*(\lambda_c)$ as λ goes to λ_c. These results will be useful for establishing the existence of an equilibrium.

Lemma 8. *Assume that* $\lambda > \lambda_c$. *Then:*

(i) The functions $\underline{s}\,(\lambda)\,,\overline{s}\,(\lambda)$ *are continuous in* λ.

(ii) Consider a sequence $\{\lambda_n\}_{n\in\mathbb{N}} > \lambda_c$ *such that* $\lim_{n\to\infty}\lambda_n = \lambda_c$. *Then* $\lim_{n\to\infty}\underline{s}\,(\lambda_n) = \lim_{n\to\infty}\overline{s}\,(\lambda_n) = s_*\,(\lambda_c)$.

Proof. See the Appendix. □

5.1.3 Equilibrium Existence and Uniqueness

This section proves Theorem 1. First, we show that if λ is large enough then the Principal best-responds by choosing a distribution, F_λ, such that the Agent's incentive constraint is slack at (λ, F_λ). Then we consider the infimum of such λ's, λ^*, and show that the incentive constraint binds at $(\lambda^*, F_{\lambda^*})$. Therefore, we can use Lemma 5 to conclude that λ^* is a best-response against F_{λ^*}. Since F_{λ^*} is a best-response to λ^*, the action profile $(\lambda^*, F_{\lambda^*})$ is an equilibrium. To prove uniqueness, we show that the incentive constraint cannot bind at (λ, F_λ) unless $\lambda = \lambda^*$. Then uniqueness follows from Lemma 5.

The following lemma shows that (IC) is slack at the Principal's best-response if λ is sufficiently large.

Lemma 9. *There exists a $\Lambda > 0$ such that (IC) evaluated at (λ, F_λ) is slack whenever $\lambda > \Lambda$.*

Proof. See the Appendix. □

Next, let

$$\lambda^* = \inf \{\lambda \,:\, \text{(IC) evaluated at } \ (\lambda, F_\lambda) \ \text{ is slack}\}\,. \tag{22}$$

The following lemma shows that (IC') binds at λ^*.

Lemma 10. *If nature chooses λ^*, then (IC) binds at $(\lambda^*, F_{\lambda^*})$.*

Proof.

First, we simplify the incentive constraint, (IC), when it is evaluated at (λ, F_λ). Recall from Lemma 7 that whenever $Z^c(\lambda, 0) < Z(\lambda, 0)$, the Principals' best-response, F_λ, is binary and is supported on $\{\underline{s}(\lambda), \overline{s}(\lambda)\}$. So, we can rewrite the incentive constraint, (IC), as

$$\underline{p}(\lambda)\, \underline{s}(\lambda)\, u(\underline{w}) + \overline{p}(\lambda)\, \overline{s}(\lambda) u(w(\lambda, \overline{s}(\lambda))) \geq c'(a^*)\ , \tag{IC'}$$

where $\underline{p}(\lambda)$ and $\overline{p}(\lambda)$ denotes the probability that $\overline{s}(\lambda)$ and $\underline{s}(\lambda)$ is realized, respectively, according to F_λ. By, (21), $\underline{p}(\lambda) = \overline{s}(\lambda)/\left(\overline{s}(\lambda) - \underline{s}(\lambda)\right)$ and $\overline{p}(\lambda) = 1 - \underline{p}(\lambda)$. Again by Lemma 7, if $Z^c(\lambda, 0) = Z(\lambda, 0)$, then F_λ is degenerate and the left-hand-side of (IC) is zero. So, by setting $\underline{p}(\lambda) = \overline{p}(\lambda) = 0$, (IC') coincides with (IC).

Suppose, by contradiction, that (IC') is slack at λ^*. It follows from Lemma 6 that $\lambda^* > \lambda_c$, for otherwise the Principal would choose the degenerate distribution and (IC') would be violated. By part (i) of Lemma 8 and continuity, there exists $\lambda < \lambda^*$ such that (IC') is also slack at λ, that is,

$$\underline{p}(\lambda)\, \underline{s}(\lambda)\, u(\underline{w}) + \overline{p}(\lambda)\, \overline{s}(\lambda)\, u(w(\lambda, \overline{s}(\lambda))) > c'(a^*)\,.$$

This contradicts the definition of λ^* given in (22).

Suppose now that (IC') is violated. First, we argue that $\lambda^* > \lambda_c$ and hence, the function $Z(\lambda^*, s)$ is non-convex. To see this, first observe that by continuity and the definition of λ^*, there exists a sequence $\{\lambda_n\}_{n\in\mathbb{N}} > \lambda^*$ such that $\lim_{n\to\infty} \lambda_n = \lambda^*$, and for all $n \in \mathbb{N}$,

$$\underline{p}(\lambda_n)\, \underline{s}(\lambda_n)\, u(\underline{w}) + \overline{p}(\lambda_n)\, \overline{s}(\lambda_n)\, u(w(\lambda_n, \overline{s}(\lambda_n))) > c'(a^*). \tag{23}$$

It must be the case that

$$\underline{s}(\lambda_n) < 0 < \overline{s}(\lambda_n), \tag{24}$$

for otherwise, $Z^c(\lambda_n, 0) = Z(\lambda_n, 0)$, so the Principal would choose the degenerate distribution and (IC') would be violated. Suppose, by contradiction, that $Z(\lambda^*, s)$ is convex in s. By Lemma 6, this implies that $\lambda^* = \lambda_c$, and the convexity of $Z(\lambda^*, s)$ in s together with the fact that $s_*(\lambda) > 0$ for all λ imply that

$$Z_2(\lambda^*, 0) < Z_2(\lambda^*, s_*(\lambda^*)).$$

By continuity and part (ii) of Lemma 8,

$$\lim_{n\to\infty} Z_2(\lambda_n, 0) = Z_2(\lambda^*, 0) \text{ and } \lim_{n\to\infty} Z_2(\lambda_n, \underline{s}(\lambda_n)) = Z_2(\lambda^*, s_*(\lambda^*)).$$

The convexity of $Z(\lambda^*, s)$ in s and the previous two displayed equations imply that $0 < \underline{s}(\lambda_n)$ for sufficiently large n. This contradicts (24), and we conclude that $\lambda^* > \lambda_c$ and $Z(\lambda^*, s)$ is non-convex in s.

If (IC') is violated at $\lambda^* (> \lambda^c)$, then by continuity and Lemma 8(i), there exists an $\varepsilon > 0$ such that (IC') is violated for all $\lambda \in [\lambda^*, \lambda^* + \varepsilon]$. This, again, contradicts the definition of λ^* in (22). □

Finally, we are in a position to prove Theorem 1.

Proof of Theorem 1.

If Nature chooses λ^*, then by Lemmas 6 and 7, the Principal's unique best response F_{λ^*} is the two-point distribution as given in (21). It remains to show that nature's best response to F_{λ^*} (or equivalently $\{\underline{s}(\lambda^*), \overline{s}(\lambda^*)\}$) is to choose λ^*. If the Principal chooses $\{\underline{s}(\lambda^*), \overline{s}(\lambda^*)\}$, then nature's problem is

$$\max_{\lambda\in\mathbb{R}_+} \left\{\underline{p}(\lambda^*)\, Z(\lambda, \underline{s}(\lambda^*)) + \overline{p}(\lambda^*)\, Z(\lambda, \overline{s}(\lambda^*))\right\}.$$

This problem is concave in λ, and the corresponding first-order condition is

$$\overline{p}(\lambda^*)\,\overline{s}(\lambda^*)\,u(w(\lambda,\overline{s}(\lambda^*))) + \underline{p}(\lambda^*)\,\underline{s}(\lambda^*)\,u(\underline{w}) = c'(a^*),$$

which is satisfied at $\lambda = \lambda^*$ by Lemma 10. Therefore, $\{\lambda^*, F_{\lambda^*}\}$ is an equilibrium for the zero-sum game described in Section 4, where λ^* is given in (22) and F_λ is given in (21).

Towards showing that this equilibrium is unique, first, recall that by Lemma 7, for any λ, the Principal's best-response F_λ is unique. Therefore, in any equilibrium, say $\{\lambda', F_{\lambda'}\}$, (IC') must bind. Moreover, by the definition of λ^* in (22), it must be the case that $\lambda' < \lambda^*$. We will show that there does not exist any $\lambda' < \lambda^*$ such that (IC') binds.

First, we show that the Principal's payoff is strictly increasing in λ on $[\lambda_c, \lambda^*]$. Pick any λ and λ' such that $\lambda_c < \lambda < \lambda' \leq \lambda^*$. Then

$$\begin{aligned} & \overline{p}(\lambda)\left[w(\lambda,\overline{s}(\lambda)) - \lambda\left[su(w(\lambda,\overline{s}(\lambda))) - c'(a^*)\right] + \overline{s}^2(\lambda)\right] \\ & +\underline{p}(\lambda)\left[\underline{w} - \lambda\left[su(\underline{w}) - c'(a^*)\right] + \underline{s}^2(\lambda)\right] \\ < \; & \overline{p}(\lambda')\left[w(\lambda',\overline{s}(\lambda')) - \lambda\left[su(w(\lambda',\overline{s}(\lambda'))) - c'(a^*)\right] + \overline{s}^2(\lambda')\right] \\ & +\underline{p}(\lambda')\left[\underline{w} - \lambda\left[su(\underline{w}) - c'(a^*)\right] + \underline{s}^2(\lambda')\right] \\ \leq \; & \overline{p}(\lambda')\left[w(\lambda',\overline{s}(\lambda')) - \lambda'\left[su(w(\lambda',\overline{s}(\lambda'))) - c'(a^*)\right] + \overline{s}^2(\lambda')\right] \\ & +\underline{p}(\lambda')\left[\underline{w} - \lambda'\left[su(\underline{w}) - c'(a^*)\right] + \underline{s}^2(\lambda')\right], \end{aligned} \tag{25}$$

where the first inequality follows from the fact that the second expression corresponds to the Principal's payoff if Nature chooses λ but the Principal uses the suboptimal information acquisition policy $\{\underline{s}(\lambda'), \overline{s}(\lambda')\}$ instead of $\{(\underline{s}(\lambda), \overline{s}(\lambda))\}$ and pays the suboptimal wage $w(\lambda', \overline{s}(\lambda'))$ instead of $w(\lambda, \overline{s}(\lambda))$. This inequality is strict because $w(\lambda', \overline{s}(\lambda')) \neq w(\lambda, \overline{s}(\lambda))$. The second inequality follows from $\lambda < \lambda'$ and the fact that (IC') is not slack if $\lambda' \leq \lambda^*$ (see 22).

Suppose, by contradiction, that there exists a $\lambda < \lambda^*$ such that (IC') binds at λ and let $\lambda' \in (\lambda, \lambda^*]$. It must be that $\lambda > \lambda_c$ for otherwise (IC') was violated. Then

$$\begin{aligned} & \overline{p}(\lambda')\left[w(\lambda',\overline{s}(\lambda')) - \lambda'\left[su(w(\lambda',\overline{s}(\lambda'))) - c'(a^*)\right] + \overline{s}^2(\lambda')\right] \\ & +\underline{p}(\lambda')\left[\underline{w} - \lambda'\left[su(\underline{w}) - c'(a^*)\right] + \underline{s}^2(\lambda')\right] \\ \leq \; & \overline{p}(\lambda)\left[w(\lambda,\overline{s}(\lambda)) - \lambda'\left[su(w(\lambda,\overline{s}(\lambda))) - c'(a^*)\right] + \overline{s}^2(\lambda)\right] \\ & +\underline{p}(\lambda)\left[\underline{w} - \lambda'\left[su(\underline{w}) - c'(a^*)\right] + \underline{s}^2(\lambda)\right] \\ = \; & \overline{p}(\lambda)\left[w(\lambda,\overline{s}(\lambda)) - \lambda\left[su(w(\lambda,\overline{s}(\lambda))) - c'(a^*)\right] + \overline{s}^2(\lambda)\right] \\ & +\underline{p}(\lambda)\left[\underline{w} - \lambda'\left[su(\underline{w}) - c'(a^*)\right] + \underline{s}^2(\lambda)\right], \end{aligned}$$

where the inequality follows from the fact that the second expression corresponds to the Principal's payoff if Nature chooses λ' but the Principal uses the suboptimal information acquisition policy $\{\underline{s}(\lambda), \overline{s}(\lambda)\}$ instead of $\{(\underline{s}(\lambda'), \overline{s}(\lambda'))\}$ and pays the suboptimal wage $w(\lambda, \overline{s}(\lambda))$ instead of $w(\lambda', \overline{s}(\lambda'))$ and the equality follows from the hypothesis that (IC') binds at λ. Finally, note that this inequality chain contradicts (25). Therefore, we conclude that there does not exist any $\lambda' < \lambda^*$ such that (IC') binds, which completes the proof. □

6 A First-Best Result

If effort was contractible, then the Principal's optimal contract would require the Agent to exert effort a^* in exchange for a wage $\underline{w}$. Of course, the Agent exerts a^* and the Principal does not acquire any information. The Principal's cost and the Agent's payoff would be $\underline{w}$ and $u(\underline{w}) - c(a^*)$, respectively. We refer to this outcome as first-best.

The following theorem describes a condition on the Agent's utility function under which a single-bonus contract can approximate the first-best outcome arbitrarily well.

Theorem 2. *Suppose that* $\lim_{w\to\infty} u(w) = \infty$ *and*

$$\lim_{w\to\infty} \frac{[u'(w)]^3}{u''(w)} [u(w)]^{-\frac{\beta-1}{\beta}} = -\infty \text{ for some } \beta > 1\,. \tag{26}$$

Then for every $\epsilon > 0$, there exists a single-bonus wage scheme and a two-point distribution that satisfy (IC) and (LL), and the Principal's expected cost is no greater than $\underline{w} + \epsilon$.

A contract that approximates the first-best outcome prescribes to acquire little information and to pay the Agent the minimum wage $\underline{w}$ with near certainty, whereas with a small probability, a large amount of information is acquired and the Agent is paid a large wage. The first-best outcome is attainable with such a contract as long as the Agent is not too risk-averse, so it is not too expensive to motivate him with a large wage that he receives with a small probability.

The condition of the theorem, (26), is satisfied if the Agent's utility function is of the form of (HARA) with parameters $\alpha > 0$, $\gamma > 1/2$, and $\beta > -\alpha\underline{w}/(1-\gamma)$. In particular, it is satisfied if the Agent has CRRA utility with coefficient less than 1/2; *i.e.*, $u(w) = w^{1-\gamma}/(1-\gamma)$ with $\gamma > 1/2$.

Proof.

To establish this result, for each $n > 0$, we construct a sequence of binary distributions

$$F_n(s) = \begin{cases} 0 & \text{if } s < -n^{-\beta} \\ \frac{n}{n+n^{-\beta}} & \text{if } s \in [-n^{-\beta}, n) \\ 1 & \text{if } s \geq n\,. \end{cases}$$

It is easy to verify that $F_n \in \mathcal{F}$, and it is a binary distribution supported on $\{-n^{\beta}, n\}$. We define a wage scheme corresponding to each F_n which pay $\underline{w}$ if $s = -n^{-\beta}$ and $w_n > \underline{w}$ if $s = n$, where w_n is chosen such that (IC) is satisfied. For any $\beta > 1$, the Principal's cost associated with choosing F_n, which is equal to its variance, converges to 0 as n goes to zero. The key part of the proof shows that when (26) is satisfied, for any target effort a^*, there exists some $\beta > 1$ such that the expected cost of providing incentives to the Agent, which is equal to $\int (w_n - \underline{w}) dF_n(n)$ also converges to 0 as $n \to \infty$. Therefore, the Principal's expected cost converges to $\underline{w}$, yielding the desired result.

Consider a sequence of probability distributions $F_n(\cdot)$ as given in (21) with $\overline{s}_n = n$ and $\underline{s}_n = -n^{-\beta}$. Let $w(s)$ denote the Agent's wage if score s is realized. For each n, define $w(\underline{s}) = \underline{w}$, and $w(\overline{s}_n)$ such that (IC) binds; that is, $w(\overline{s}_n)$ satisfies

$$\left(1 + \frac{\underline{s}_n}{\overline{s}_n - \underline{s}_n}\right) \underline{s}_n u(\underline{w}) - \frac{\underline{s}_n}{\overline{s}_n - \underline{s}_n} \overline{s}_n u(w(\overline{s}_n)) = c'(a^*)$$
$$\Rightarrow w(\overline{s}_n) = u^{-1}\left(u(\underline{w}) - \frac{\overline{s}_n - \underline{s}_n}{\overline{s}_n \underline{s}_n} c'(a^*)\right)\,,$$

where we used that $\Pr(\overline{s}_n) = -\underline{s}_n/(\overline{s}_n - \underline{s}_n)$. Observe that $w(\overline{s}_n) > \underline{w}$, so (LL) is satisfied for all n. The Principal's expected cost is equal to

$$\begin{aligned}\Pi_n &= \frac{\overline{s}_n}{\overline{s}_n - \underline{s}_n}\left[\underline{w} + \underline{s}_n^2\right] - \frac{\underline{s}_n}{\overline{s}_n - \underline{s}_n}\left[w(\overline{s}_n) + \overline{s}_n^2\right] \\ &= \frac{\overline{s}_n}{\overline{s}_n - \underline{s}_n}\underline{w} - \frac{\underline{s}_n}{\overline{s}_n - \underline{s}_n} u^{-1}\left(u(\underline{w}) - \frac{\overline{s}_n - \underline{s}_n}{\overline{s}_n \underline{s}_n} c'(a^*)\right) + \frac{\overline{s}_n \underline{s}_n^2 - \underline{s}_n \overline{s}_n^2}{\overline{s}_n - \underline{s}_n}\,.\end{aligned}$$

We will show that $\Pi_n \to \underline{w}$ as $n \to \infty$ for some $\beta > 1$. First, for all $\beta > 1$,

$$\lim_{n\to\infty} \frac{\overline{s}_n \underline{s}_n^2 - \underline{s}_n \overline{s}_n^2}{\overline{s}_n - \underline{s}_n} = \lim_{n\to\infty} \frac{n^{1-2\beta} + n^{2-\beta}}{n + n^{-\beta}} \leq \lim_{n\to\infty}\left(n^{-2\beta} + n^{1-\beta}\right) = 0 \tag{27}$$

Next, note that for every $n \geq 1$,

$$0 \leq -\frac{\underline{s}_n}{\overline{s}_n - \underline{s}_n} u^{-1}\left(u(\underline{w}) - \frac{\overline{s}_n - \underline{s}_n}{\overline{s}_n \underline{s}_n} c'(a^*)\right) = \frac{n^{-\beta}}{n + n^{-\beta}} u^{-1}\left(u(\underline{w}) + \frac{n + n^{-\beta}}{n^{1-\beta}} c'(a^*)\right)$$
$$\leq \frac{u^{-1}\left(v + n^{\beta} c'(a^*)\right)}{n^{\beta+1} + 1}, \tag{28}$$

where we let $v = u(\underline{w}) + c'(a^*)$. Observe that both the nominator, and the denominator diverge to ∞ as $n \to \infty$. Applying L'Hospital's rule, we have that

$$\lim_{n\to\infty} \frac{u^{-1}\left(v + n^{\beta} c'(a^*)\right)}{n^{\beta+1} + 1} = \lim_{n\to\infty} \frac{\frac{\beta n^{\beta-1} c'(a^*)}{u'\left(u^{-1}\left(v + n^{\beta} c'(a^*)\right)\right)}}{(\beta+1) n^{\beta}} \leq c'(a) \lim_{n\to\infty} \frac{\frac{1}{u'\left(u^{-1}\left(v + n^{\beta} c'(a^*)\right)\right)}}{n}.$$

If $\lim_{w\to\infty} u'(w) > 0$ (*e.g.*, if $u(\cdot)$ is affine), then the last limit equals 0, so $\lim_{n\to\infty} \Pi_n = \underline{w}$, and the proof is complete. Suppose that this is not the case, so both the nominator and the denominator diverge to ∞ as $n \to \infty$. Applying L'Hospital's rule (again), we have that

$$c'(a) \lim_{n\to\infty} \frac{\frac{1}{u'\left(u^{-1}\left(v + n^{\beta} c'(a^*)\right)\right)}}{n} = \beta [c'(a^*)]^2 \lim_{n\to\infty} \frac{-u''\left(u^{-1}\left(v + n^{\beta} c'(a^*)\right)\right)}{\left[u'\left(u^{-1}\left(v + n^{\beta} c'(a^*)\right)\right)\right]^3} n^{\beta-1}.$$

Letting $w = u^{-1}\left(v + n^{\beta} c'(a^*)\right)$, the last expression can be re-written as

$$\beta [c'(a^*)]^2 \lim_{w\to\infty} \frac{-u''(w)}{[u'(w)]^3} \left[\frac{u(w) - v}{c'(a^*)}\right]^{\frac{\beta-1}{\beta}} \leq \beta [c'(a^*)]^{\frac{\beta+1}{\beta}} \lim_{w\to\infty} \frac{-u''(w)}{[u'(w)]^3} [u(w)]^{\frac{\beta-1}{\beta}},$$

where the inequality follows because $u(w) > v$ and $\beta \geq 1$. By assumption, there exists some $\beta > 1$ such that the last limit is equal to 0. Therefore $\lim_{n\to\infty} \Pi_n = \underline{w}$, and so for any $\epsilon > 0$, there exists some N such that $\Pi_n < \underline{w} + \epsilon$ for all $n > N$. Finally, note that by construction, $\underline{s}_n$, $\overline{s}_n$, $w(\underline{s}) = \underline{w}$, and $w(\overline{s}_n)$ satisfy (IC) and (LL). □

7 Validating the First-Order Approach

Throughout the analysis, we have considered a relaxed problem, in which the Principal restricts attention to discouraging local downward deviations from the target effort a^*. In this section, we consider an arbitrary binary distribution over scores and a wage scheme that satisfies the relaxed incentive compatibility constraint given in (5), and we provide sufficient conditions such that this contract also satisfies the global incentive compatibility constraint given in (2).[13]

In what follows, we fix a binary distribution over scores, F, and a wage scheme $\left\{\widetilde{W}(s)\right\}_s$. First, we compute the probability of each score conditional on any deviation, $a\,(\neq a^*)$. Second, we use these conditional probabilities to express the Agent's global incentive constraint. Finally, we provide conditions under which the first-order condition at a^* implies that this global incentive constraint is satisfied. To this end, let $\{\underline{s}, \overline{s}\}$ be the support of the distribution, where $\underline{s} < \overline{s}$. This distribution is implemented by the stopping time $\tau = \inf\{t : s_t \notin (\underline{s}, \overline{s})\}$, where $ds_t = (a - a^*)dt + dB_t$. Let $p(a)$ denote the probability that $\overline{s}$ is realized given the Agent's effort a. Using Ito's Lemma, it is not hard to show that

$$p(a) = \frac{e^{-2(a-a^*)\underline{s}} - 1}{e^{-2(a-a^*)\underline{s}} - e^{-2(a-a^*)\overline{s}}}. \tag{29}$$

[14] Then the Agent's problem is

$$\max_a u\left(\widetilde{W}\left(\underline{s}\right)\right) + p\left(a\right)\left[u\left(\widetilde{W}\left(\overline{s}\right)\right) - \widetilde{W}\left(\underline{s}\right)\right] - c\left(a\right). \tag{30}$$

The first-order condition at a^* is

$$p'\left(a^*\right) u(w(\overline{s})) - u(w(\underline{s})) = c'(a^*),$$

that is, wages must satisfy $u(w(\overline{s})) - u(w(\underline{s})) = c'(a^*)/p'(a^*)$. Plugging this into (30), the Agent's global incentive compatibility constraint can be expressed as

$$a^* \in \arg\max_{a \geq 0} \left\{ u(w(\underline{s})) + p(a)\frac{c'(a^*)}{p'(a^*)} - c(a) \right\}.$$

This constraint is satisfied if the maximand is single-peaked at a^*. Note that the maximand's derivative is $p'\left(a\right)\left[c'\left(a^*\right)/p'\left(a^*\right)\right] - c'\left(a\right)$. So, the first-order approach is valid if

$$\frac{p'(a)}{p'(a^*)} \geq \frac{c'(a)}{c'(a^*)} \quad \text{if and only if } a \leq a^*. \tag{31}$$

[13]In the canonical Principal-Agent model (*e.g.,* Holmström (1979)), to ensure that the first-order approach is valid, it is typically assumed that either that the transition probability function which maps each effort level into contractible output is convex in effort, or conditions are imposed on that transition probability function and the Agent's utility function; see Bolton and Dewatripont (2005) and Jewitt (1988) for details. In our setting, this distribution is endogenous, and hence we impose conditions on the Agent's effort cost function.

[14]To derive (29), fix an effort level $a \neq a^*$, and let $\overline{p}(s)$ denote the probability that $s_\tau = \overline{s}$ given the current score s. Applying Ito's lemma on $ds_t = (a - a^*)dt + dB_t$, it follows that $\overline{p}$ satisfies $0 = 2(a - a^*)\overline{p}'(s) + \overline{p}''(s)$ subject to the boundary conditions $\overline{p}(\overline{s}) = 1$ and $\overline{p}(\underline{s}) = 0$. This boundary value problem can be solved analytically, and evaluating its solution at $s_0 = 0$ yields (29). Moreover, it follows from L'Hospital's rule that $\left(1 - e^{-2(a-a^*)\overline{s}}\right) / \left(e^{-2(a-a^*)\underline{s}} - e^{-2(a-a^*)\overline{s}}\right) \to \overline{s}/(\overline{s} - \underline{s})$ as $a \to a^*$, and so (29) corresponds to (21) when $a = a^*$.

The following proposition gives sufficient conditions for this to be the case.

Proposition 1. *Consider a sequence of effort cost functions* $\{c_k\}_{k\in\mathbb{N}}$ *such that* $c_k'(a^*) = d$ *for all* k*, where* $d > 0$ *is some constant. In addition,*

(i) for all $a < a^*$, $c_k'(a)$ *is decreasing in* k *and* $\lim_{k\to\infty} c_k'(a) = 0$,

(ii) for all $a > a^*$, $c_k'(a)$ *is increasing in* k *and* $\lim_{k\to\infty} c_k'(a) = \infty$,

(iii) and $c_k''(a^*)$ *is increasing in* k*, and* $\lim_{k\to\infty} c_k''(a^*) = \infty$.

Then there exists a $K \in \mathbb{N}$ *such that the first-order approach is valid if* $c(a) = c_k(a)$ *whenever* $k > K$.

The first-order approach is valid as long as the Agent's effort cost function is sufficiently convex at the target effort level a^*, sufficiently flat at effort levels below a^*, and sufficiently steep at effort levels above a^*. For example, it is valid if $c'(a) = a^k$ and $a^* = 1$ for a sufficiently large k.

Proof.

First, fix an effort level $a < a^*$. By condition (i), there exists $K_a \in \mathbb{N}$ such that

$$\frac{p'(a)}{p'(a^*)} \geq \frac{c'_{K_a}(a)}{d}.$$

Furthermore, by continuity, there exists $\varepsilon_a\,(< a^* - a)$ such that the previous inequality is satisfied for all $\widetilde{a} \in (\max\{0, a - \varepsilon_a\}, a + \varepsilon_a) = N_a$. Since $c_k'(\widetilde{a})$ is decreasing in k by condition (i), it follows that for all $\widetilde{a} \in N_a$ and $k \geq K_a$,

$$\frac{p'(\tilde{a})}{p'(a^*)} \geq \frac{c_k'(\tilde{a})}{d}. \tag{32}$$

Second, fix an effort level $a > a^*$. By conditions (ii), there exists $K_a' \in \mathbb{N}$ such that

$$\frac{p'(a)}{p'(a^*)} \leq \frac{c'_{K_a'}(a)}{d}.$$

Furthermore, by continuity, there exists $\varepsilon_a\,(< a - a^*)$ such that the previous inequality is satisfied for all $\widetilde{a} \in (a - \varepsilon_a, a + \varepsilon_a) = N_a$. Since $c_k'(\widetilde{a})$ is increasing in k by condition (ii), it follows that for all $\widetilde{a} \in N_a$ and $k \geq K_a'$,

$$\frac{p'(\tilde{a})}{p'(a^*)} \leq \frac{c_k'(\tilde{a})}{d}. \tag{33}$$

Now, consider $a = a^*$. By condition (iii) and noting that $p''(a^*)/p'(a^*) = (2/3)(\underline{s} + \overline{s})$ is finite, there exists K_{a^*} such that the local second-order condition is satisfied, that is,

$$\frac{p''(a^*)}{p'(a^*)}d < c''_{K_{a^*}}(a^*)\,.$$

Therefore, there exists $\varepsilon_{a^*} > 0$, such that if $\widetilde{a} \in (a - \varepsilon_{a^*}, a + \varepsilon_{a^*}) = N_{a^*}$, then (31) is satisfied with $c = c_{K_{a^*}}$. By the monotonicity properties in conditions (i) and (ii), for all $\widetilde{a} \in N_{a^*}$ and $k \geq K_{a^*}$,

$$\begin{aligned} \frac{p'(\tilde{a})}{p'(a^*)} &> \frac{c'_k(\widetilde{a})}{d} \text{ if } \widetilde{a} < a^*, \\ \frac{p'(\tilde{a})}{p'(a^*)} &< \frac{c'_k(\widetilde{a})}{d} \text{ if } \widetilde{a} > a^*. \end{aligned} \tag{34}$$

Using L'Hospital's rule, one can show that for any a^*, $\lim_{a\to\infty} p'(a) = 0$. Therefore, there exists some B such that $p'(a)/p'(a^*) < B$ for all $a \geq 0$, and hence for any $A > a^*$, there exists some K_A such that

$$\frac{p'(a)}{p'(a^*)} \leq \frac{c'_k(a)}{c'_k(a^*)} \quad \text{for all } a > A \text{ and } k \geq K_A\,.$$

Fix some $A > a^*$ and note that

$$[0, A] = \cup_{a\in[0,A]} N_a.$$

Since $[0, A]$ is compact and N_a is open for all $a \in [0, A]$ there exists $a_1, ..., a_m$ such that

$$[0, A] = \cup_{j\in\{1,...,m\}} N_{a_j}. \tag{35}$$

Now, let us define

$$K = \max\{K_{a_1}, ..., K_{a_m}, K_A\} \tag{36}$$

and let us consider c_k such that $k > K$. We show that the first-order approach is valid if the Agent's effort cost is given by c_k. To be more specific, we show that c_k satisfies (31). To this end, suppose first that $a < a^*$. By (35), there is $j \in \{1, ..., m\}$ such that $a \in N_j$. Note that either $a_j < a^*$ or $a_j = a^*$. If $a_j < a^*$, then a satisfies (32) because $k > K > K_{a_j}$ by (36). If $a_j = a^*$ then a satisfies the first line of (34) because $k > K > K_{a_j}$. Suppose now that $a > a^*$. Again, by (35), there is $j \in \{1, ..., m\}$ such that $a \in N_j$. Note that either $a_j > a^*$, or $a_j = a^*$. If $a_j > a^*$, then a satisfies (33) because $k > K > K_{a_j}$ by (36). If $a_j = a^*$ then a satisfies the second line of (34) because $k > K > K_{a_j}$. Finally, (31) is satisfied for all $a > A$ since $K \geq K_A$ and the proof is complete. □

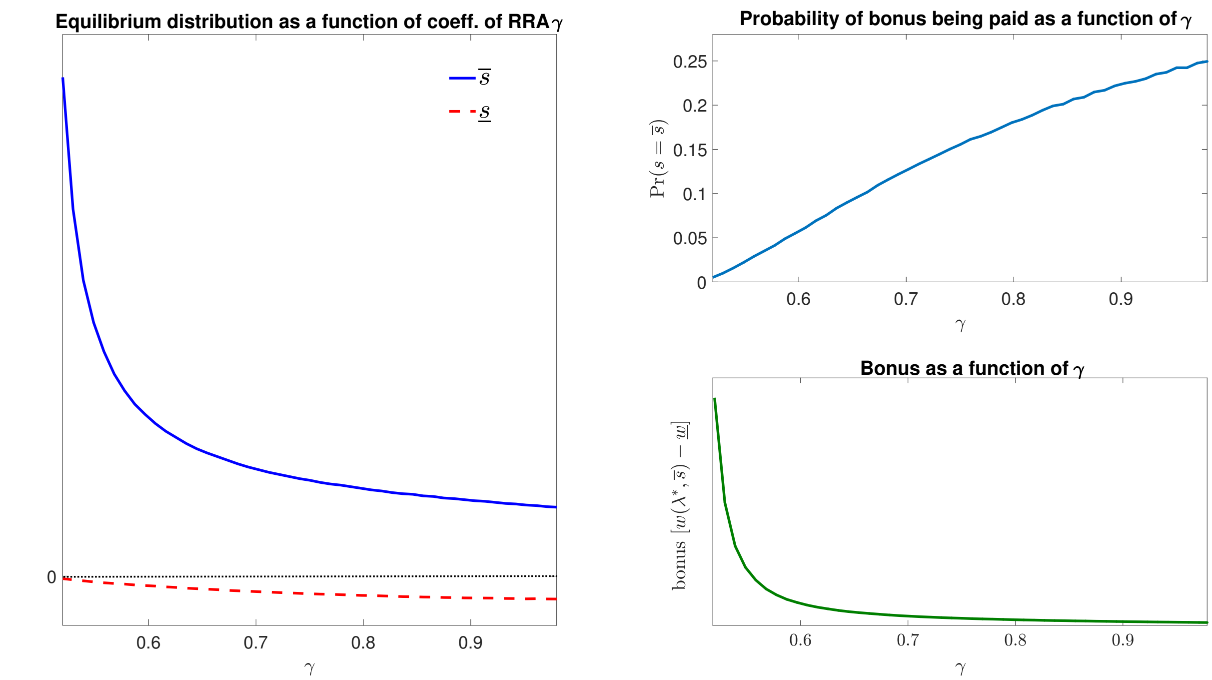

Figure 2: Comparative statics of the optimal contract as the (constant) coefficient of relative risk aversion γ varies.

8 Comparative Statics

In this section, we use simulations to investigate how the optimal contract depends on the parameters of the problem.

Figure 2 provides comparative statics when the Agent's utility function exhibits CRRA, and so it is of the form of

$$u(w) = \frac{w^{1-\gamma}}{1-\gamma},$$

and we vary the coefficient of relative risk aversion γ from $1/2$ to 1, while setting $\underline{w} = 0.1$ and $c'(a^*) = 1$. The left panel illustrates the scores in the support of the equilibrium distribution, $\overline{s}$ and $\underline{s}$, while the right panels illustrate the size of the bonus $(w(\lambda^*, \overline{s}) - \underline{w})$ and the probability that it is paid $(\overline{p}(\lambda^*))$ as a function of γ. As this figure illustrates, if γ is close to a half and the Agent is moderately risk-averse then the optimal specifies an $\underline{s}$ close to zero and a large $\overline{s}$. This implies that the Agent only receives the bonus if the acquired information is overwhelmingly favorable. Of course, this event occurs with only a small probability as shown on the top-right panel of Figure 2. Therefore, in order to motivate the Agent, the size of the bonus must be large confirmed by the right-bottom panel. Note that the equilibrium contract in this case is similar to the ones used in the proof of Theorem 2: the Agent receives a large bonus with a small probability if she exerts the target effort. Recall that this theorem implies that if $\gamma < 1/2$, then it is possible to approximate the first-best outcome arbitrarily closely

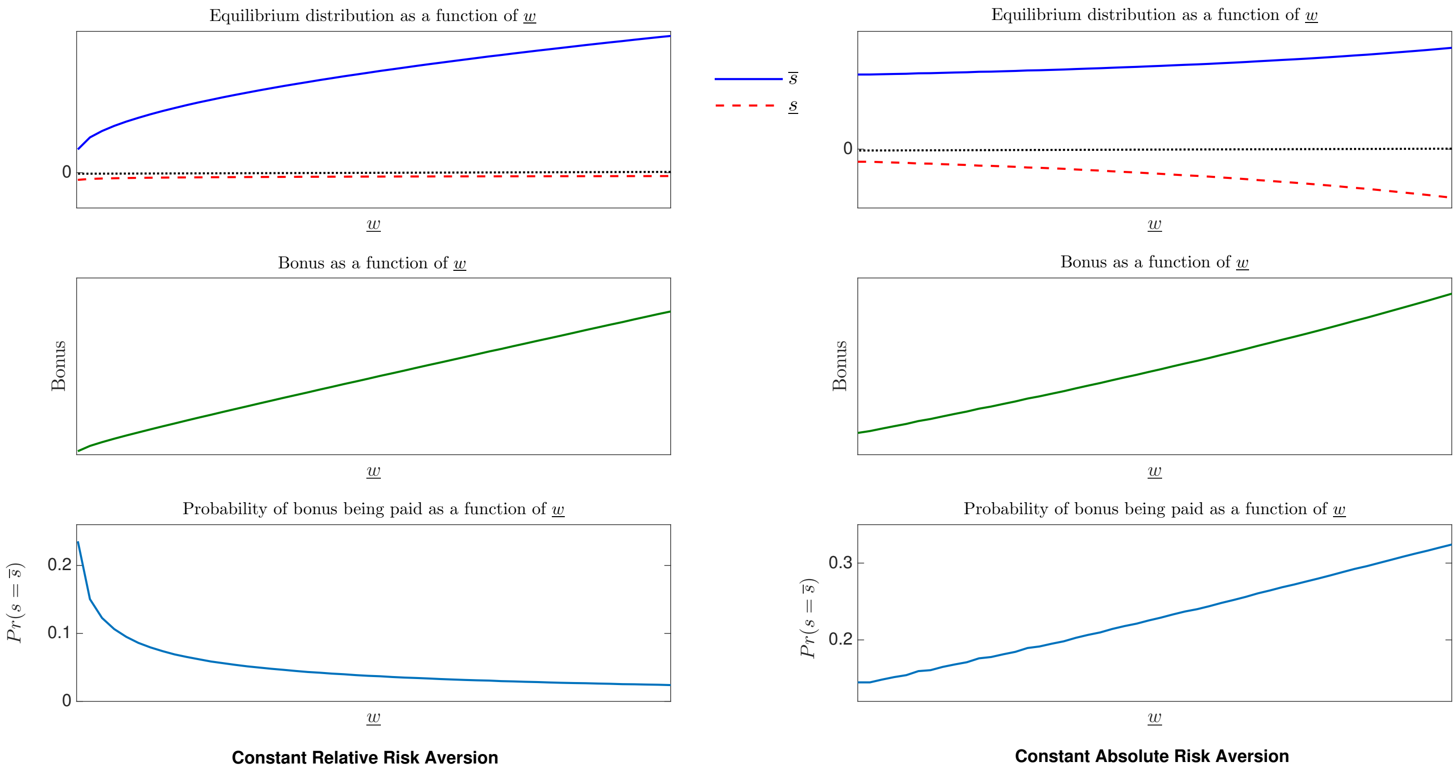

Figure 3: Comparative statics of the optimal contract as the minimum wage $\underline{w}$ varies. The left (right) panel illustrates the case in which the Agent's utility exhibits CRRA, and constant relative (absolute) risk aversion (left panel).

using such single-bonus wage schemes.

In contrast, if γ is close to one and the Agent is very risk averse then both $\underline{s}$ and $\overline{s}$ are relatively small (see the left panel of Figure 2). This implies that the Agent receives a bonus with high probability as shown on the top-right panel. Since the Agent receives the bonus frequently, its size is small. The fact that the optimal contract specifies small rewards with large probability if γ is large is not surprising given that a very risk-averse Agent values income-smoothing more. Note that between the two extreme values of γ, all the functions are monotone: $\underline{s}$, $\overline{s}$, and the bonus are decreasing and the probability of the bonus increases. Simulations indicate that these comparative statics are similar when the Agent's utility function exhibits CARA, and we vary the coefficient of absolute risk aversion.

Figure 3 illustrates how the scores in the support of the equilibrium distribution, $\overline{s}$ and $\underline{s}$, the size of the bonus, and the probability that it is paid vary with the minimum wage $\underline{w}$, when the Agent's utility exhibits CRRA or CARA. In the former case, we set $\gamma = 0.9$ and $c'(a^*) = 1$, and vary $\underline{w}$ from 0 to 5. In the latter case, we set the coefficient of absolute risk aversion to 1 and $c'(a^*) = 1$, and vary $\underline{w}$ from -5 to 0. In both cases, the size of the bonus increases in the minimum wage. The reason is that, as $\underline{w}$ increases, the marginal utility of the Agent for a given given bonus decreases. Therefore, in order to incentivize the Agent to exert the target effort, the Principal must increase the size of the bonus. The comparative statics pertaining to the optimal information acquisition strategies appear to be quite different in the two cases examined. In particular, if the Agent's utility exhibits CRRA then the probability of paying the bonus is decreasing in $\underline{w}$, whereas this probability is increasing if the case of CARA utility. The reason is that in the former case, as $\underline{w}$ increases the Agent's become less and less risk-averse regarding gambles involving transfers above $\underline{w}$. As a consequence, the optimal contract specifies the familiar small-probability, large-bonus wage scheme. In the case of CARA, such an effect is not present.

9 Discussion

We analyze a contracting problem under moral hazard in which the Principal designs both the Agent's wage scheme, and the underlying performance measure. In our model, a performance measure is a strategy for sequentially acquiring signals that are informative of the Agent's costly effort, and a wage scheme specifies the Agent's remuneration conditional on the acquired signals. Under a pair of conditions on the Agent's utility function, and provided that the first-order approach is valid, we show that a single-bonus contract is optimal; *i.e.,* the Principal chooses a two-point distribution over scores and a binary wage scheme. These conditions are satisfied if, for example, the Agent's utility exhibits CARA, or CRRA with coefficient greater than half. Under an alternative condition on the Agent's utility, which is satisfied if, for instance, it exhibits CRRA with coefficient less than half, we show that the Principal can approximate the first-best outcome arbitrarily closely with a single-bonus contract.

Throughout the paper, we assumed that the Agent has limited liability but does not make a participation decision. It is not hard to incorporate a participation constraint into the Principal's optimization problem and show that the optimal contract still involves binary wages. We chose not to do so because this additional constraint has little to do with the main argument of our analysis and involves heavy notational burden. Our results are also valid if the Agent's effort is binary. In this case, there is only a single incentive constraint and one does not need to consider a relaxed problem. As a consequence, the optimality of a single-bonus contract does not require those conditions on the Agent's effort cost which we imposed to validate the first-order approach.

We have considered a particular information acquisition mechanism, which is equivalent to the Principal choosing any zero-mean distribution over scores at a cost equal its variance. Alternatively, we could have started with an information design problem in which the Principal chooses a distribution over scores $F \in \mathcal{F}$ at some cost. Our main theorem holds as long as this cost is a general convex moment, that is, it can be expressed as $\mathbb{E}_F[\varphi(s)]$ for some function φ with $\varphi'' > 0$ and $\varphi''' \geq 0$.[15] Recall that in our model, by choosing a distribution corresponding to the target effort level, the Principal is implicitly choosing a distribution corresponding to every other effort level. A drawback of this alternative approach is that it is ambiguous how the distributions corresponding to different effort levels ought to be linked, and so absent additional assumptions, it is impossible to validate the first-order approach.

Prior to acquiring (costly) information, the Principal is completely oblivious to the Agent's effort in our model. This assumption is helpful for highlighting the mechanism of the model, and it provides tractability. However, in some settings, the Principal observes costless information pertaining to the Agent's effort, which she can augment by acquiring additional signals. For example, public companies are obligated to report various accounting measures such as revenue and operating profits, which are likely to be informative of some employees' actions. Such costless information can be incorporated into our model if, prior to acquiring (costly) information, the Principal observes a signal that is correlated with the Agent's effort. Let s_0 denote the score associated with the costless signal.[16] Provided that the first-order approach is valid, the Principal's problem can be expressed as choosing a family of distributions over scores, one for each s_0, with mean equal to s_0, and a wage scheme conditional on the realized score. Using similar techniques as in Sections 3 and 4, one can show that an optimal contract corresponds to an equilibrium of a zero-sum game. Unfortunately, existence of an equilibrium is difficult to establish there. However, under the conditions of Theorem 1, *if* an equilibrium exists, then the optimal contract takes the following form: If s_0 lies below a lower threshold, say s_l, or above an upper threshold, say s_h, then the Principal optimally chooses a distribution over scores that is degenerate at s_0 (which corresponds to acquiring no costly information), and pays the Agent the minimum wage $\underline{w}$, or a larger wage that depends on s_0, respectively. On the other hand, if s_0 lies between s_l and s_h, then the Principal chooses a two-point distribution that assigns positive probability only on s_l and s_h, and pays the Agent either $\underline{w}$ or a larger wage, depending on which score is realized. Interestingly, this wage scheme resembles what Murphy (1999) (Figure 5) and Jensen (2001) argue is a *typical executive incentive plan.*

[15] In our model, $\varphi(s) = s^2$. The shape of φ is only used in the proof of Lemma 6, which requires that φ is strictly convex, and $\varphi''' \geq 0$ so that the threshold λ_c is unique and well-defined.

[16] If the costless signal $x_0 \sim H(\cdot|a)$ for some CDF $H(\cdot|a)$, then the score $s_0 = dH_a(x_0|a^*)/dH(x_0|a^*)$.

References

Aumann, R.J. and Perles, M., 1965. A variational problem arising in economics. *Journal of Mathematical Analysis and Applications*, 11, pp.488-503.

Baiman, S. and Demski, J.S., 1980. Economically optimal performance evaluation and control systems. *Journal of Accounting Research*, pp.184-220.

Barron, D., Georgiadis, G. and Swinkels, J., 2016. Optimal Contracts with a Risk-Taking Agent. Working Paper.

Boleslavsky, R. and Kim, K., 2017. Bayesian Persuasion and Moral Hazard. Working Paper.

Bolton, P. and Dewatripont, M., 2005. Contract Theory. MIT Press.

Buckingham, M. and Goodall, A., 2015. Reinventing Performance Management. *Harvard Business Review*, 93(4), pp.40-50.

Carroll, G., 2015. Robustness and linear contracts. *American Economic Review*, 105(2), pp. 536-563.

Datar, S., Kulp, S.C. and Lambert, R.A., 2001. Balancing Performance Measures. *Journal of Accounting Research*, 39(1), pp.75-92.

Dye, R.A., 1986. Optimal monitoring policies in agencies. *RAND Journal of Economics*, pp.339-350.

Feltham, G.A. and Xie, J., 1994. Performance Measure Congruity and Diversity in Multi-task Principal/agent Relations. *Accounting Review*, pp.429-453.

Grossman, S. and Hart, O.D., 1983. An Analysis of the Principal-Agent Problem. *Econometrica,* 51(1), pp.7-45.

Hart, O.D. and Holmström, B., 1986. *The Theory of Contracts*. Department of Economics, Massachusetts Institute of Technology.

Hoffmann, F., Inderst, R. and Opp, M.M., 2017. Only time will tell: A theory of deferred compensation and its regulation. Working Paper.

Holmström, B., 1979. Moral Hazard and Observability. *Bell Journal of Economics*, pp.74-91.

Holmström, B., 2016. Nobel Prize Lecture. Retrieved from Nobelprize.org.

Holmström, B. and Milgrom, P., 1987. Aggregation and linearity in the provision of intertemporal incentives. *Econometrica*, pp. 303-328.

Jensen, M.C., 2001. Corporate Budgeting Is Broken, Let's Fix It. *Harvard Business Review*, 79(10), pp.94-101.

Jewitt, I., 1988. Justifying the First-order Approach to Principal-agent Problems. *Econometrica*, pp. 1177-1190.

Jewitt, I., Kadan, O. and Swinkels, J.M., 2008. Moral Hazard with Bounded Payments. *Journal of Economic Theory*, 143(1), pp.59-82.

Kamenica, E. and Gentzkow, M., 2011. Bayesian Persuasion. *American Economic Review*, 101(6), pp. 2590-2615.

Kim, S.K. and Suh, Y.S., 1992. Conditional monitoring policy under moral hazard. *Management Science*, 38(8), pp.1106-1120.

Levin, J., 2003. Relational Incentive Contracts. American Economic Review, pp.835-857.

Li, A. and Yang, M., 2017. Optimal Incentive Contract with Endogenous Monitoring Technology. Working Paper.

Mauboussin M.L., 2012. The True Measures of Success. *Harvard Business Review.*

Mirrlees, J.A., 1976. The optimal structure of incentives and authority within an organization. *Bell Journal of Economics*, pp. 105-131.

Murphy, K.J., 1999. Executive Compensation. *Handbook of Labor Economics*, 3, pp. 2485-2563.

Palomino, F. and Prat, A., 2003. Risk taking and optimal contracts for money managers. *RAND Journal of Economics*, pp. 113-137.

Root, D.H., 1969. The existence of certain stopping times on Brownian motion. *The Annals of Mathematical Statistics*, 40(2), pp.715-718.

Rost, H., 1976. Skorokhod stopping times of minimal variance. In Séminaire de Probabilités X Université de Strasbourg (pp. 194-208). Springer, Berlin, Heidelberg.

Rayo, L. and Segal, I., 2010. Optimal information disclosure. *Journal of Political Economy*, 118(5), pp.949-987.

Townsend, R.M., 1979. Optimal contracts and competitive markets with costly state verification. *Journal of Economic Theory*, 21(2), pp.265-293.

von Neumann, J.,1928. Zur Theorie der Geschaftsspiele. *Mathematische Annalen*, 100, pp. 295-320.

WorldatWork and Deloitte Consulting LLP, 2014. Incentive Pay Practices Survey: Publicly Traded Companies.

Young, R.A., 1986. A note on" Economically optimal performance evaluation and control systems": The optimality of two-tailed investigations. *Journal of Accounting Research*, pp.231-240.

Proof of Lemma 1.

Fix an arbitrary stopping rule τ, and assume that the wage scheme $W(\omega_\tau)$ solves (1) subject to (3) and (5). Note that $B_\tau(\omega_\tau) = \omega_\tau - a^*\tau$. Suppose that there exist at least two paths ω_τ^1 and ω_τ^2 such that $B_\tau(\omega_\tau^1) = B_\tau(\omega_\tau^2)$ and $W(\omega_\tau^1) \neq W(\omega_\tau^2)$. Define a new wage scheme

$$W^{(2)}(\omega_\tau) = \begin{cases} \mathbb{E}_{a^*}\left[W(\omega_\tau)|\omega_\tau \in \{\omega_\tau^1, \omega_\tau^2\}\right] & \text{if } \omega_\tau \in \{\omega_\tau^1, \omega_\tau^2\} \\ W(\omega_\tau) & \text{otherwise.} \end{cases}$$

By construction, this wage scheme bears the same expected cost to the Principal and satisfies (3). Notice that

$$\begin{aligned}
&\mathbb{E}_{a^*}\left[u(W^{(2)}(\omega_\tau))B_\tau(\omega_\tau)\right] \\
=&\mathbb{E}_{a^*}\left[u(W^{(2)}(\omega_\tau))B_\tau(\omega_\tau)|\omega_\tau \in \{\omega_\tau^1, \omega_\tau^2\}\right] + \mathbb{E}_{a^*}\left[u(W^{(2)}(\omega_\tau))B_\tau(\omega_\tau)|\omega_\tau \notin \{\omega_\tau^1, \omega_\tau^2\}\right] \\
=&u\left(\mathbb{E}_{a^*}\left[W(\omega_\tau)|\omega_\tau \in \{\omega_\tau^1, \omega_\tau^2\}\right]\right) B_\tau(\omega_\tau^1) + \mathbb{E}_{a^*}\left[u(W(\omega_\tau))B_\tau(\omega_\tau)|\omega_\tau \notin \{\omega_\tau^1, \omega_\tau^2\}\right] \\
\geq&\mathbb{E}_{a^*}\left[u(W(\omega_\tau))|\omega_\tau \in \{\omega_\tau^1, \omega_\tau^2\}\right] B_\tau(\omega_\tau^1) + \mathbb{E}_{a^*}\left[u(W(\omega_\tau))B_\tau(\omega_\tau)|\omega_\tau \notin \{\omega_\tau^1, \omega_\tau^2\}\right] \\
=&\mathbb{E}_{a^*}\left[u(W(\omega_\tau))B_\tau(\omega_\tau)\right] \geq c'(a^*),
\end{aligned}$$

where the first equality follows by expressing the expectation conditionally on whether $\omega_\tau \in \{\omega_\tau^1, \omega_\tau^2\}$, the second equality follows from the definition of $W^{(2)}(\omega_\tau)$, the first inequality follows from Jensen's inequality, the third equality follows by collecting terms, and the second inequality follows because by assumption, $W(\omega_\tau)$ satisfies (5). Therefore, $W^{(2)}(\omega_\tau)$ also solves (1) subject to (3) and (5).

Repeat this process iteratively until there exist no two paths, say ω'_τ and ω''_τ, such that $B_\tau(\omega'_\tau) = B_\tau(\omega''_\tau)$ and $W^{(n)}(\omega'_\tau) \neq W^{(n)}(\omega''_\tau)$. Then the wage scheme $W^{(n)}(\omega_\tau)$ depends only on $B_\tau(\omega_\tau)$, and it solves (1) subject to (3) and (5). □

Proof of Lemma 3.

Part 1: Optimality & uniqueness

Suppose that there exists a $\lambda^* > 0$ such that $w(\lambda^*, s)$ satisfies (IC) with equality. By construction, $w(\lambda^*, s)$ also satisfies (LL) for any λ and s.

We claim that this wage scheme uniquely solves (6) subject to (IC) and (LL). Towards a contradiction, assume that there exists some $\tilde{w}(\cdot)$ which differs from $w(\lambda^*, \cdot)$ on a set of positive measure, satisfies (IC)-(LL), and bears a weakly lower expected cost to the Principal than $w(\lambda^*, \cdot)$. Define $w^\epsilon(\cdot)$ by

$$u(w^\epsilon(s)) = (1-\epsilon)u(w(\lambda^*, s)) + \epsilon u(\tilde{w}(s))$$

for all s. This is the certainty equivalently of a $(1-\epsilon, \epsilon)$ lottery between $w(\lambda^*, \cdot)$ and $\tilde{w}(\cdot)$. Note that

$$\frac{\partial w^\epsilon(s)}{\partial \epsilon} = \frac{1}{u'(w^\epsilon(s))}\left[u(\tilde{w}(s)) - u(w(\lambda^*, s))\right] \text{ and } \frac{\partial^2 w^\epsilon(s)}{\partial \epsilon^2} = -\frac{u''(w^\epsilon(s))}{\left[u'(w^\epsilon(s))\right]^3}\left[u(\tilde{w}(s)) - u(w(\lambda^*, s))\right]^2 \geq 0\,,$$

where the inequality is strict if $w(\lambda^*, s) \neq \tilde{w}(s)$. Therefore, the Principal's expected cost associated with the contract $w^\epsilon(\cdot)$, $\Pi(w^\epsilon) = \int w^\epsilon(s)dF(s)$ is strictly convex in ϵ as $\tilde{w}(s)$ and $w(\lambda^*, \cdot)$ differ on a set of positive measure. Since $\Pi(w^\epsilon) \leq \int w(\lambda^*, s)dF(s)$ by assumption, it must be the case that

$$\left.\frac{\partial \Pi(w^\epsilon)}{\partial \epsilon}\right|_{\epsilon=0} < 0\,. \tag{37}$$

We will show that

$$\frac{1}{u'(w(\lambda^*, s))}\left[u(\tilde{w}(s)) - u(w(\lambda^*, s))\right] \geq \lambda^* s\left[u(\tilde{w}(s)) - u(w(\lambda^*, s))\right]$$

for all s. This inequality holds trivially for all $s > s_*(\lambda^*)$ since $1/u'(w(\lambda^*, s)) = \lambda^* s$ for such s. If $s \leq s_*(\lambda^*)$, then $w(\lambda^*, s) = \underline{w}$, and the desired inequality follows from the facts that $u(\tilde{w}(s)) - u(w(\lambda^*, s)) \geq 0$ (as $\tilde{w}(\cdot)$ satisfies (LL)) and

$$\frac{1}{u'(w(\lambda^*, s))} = \frac{1}{u'(\underline{w})} \geq \lambda^* s\,.$$

Therefore, we have

$$\begin{aligned}
\left.\frac{\partial \Pi(w^\epsilon)}{\partial \epsilon}\right|_{\epsilon=0} &= \int \frac{\partial w^\epsilon(s)}{\partial \epsilon} dF(s) \\
&= \int \frac{1}{u'(w^\epsilon(s))}\left[u(\tilde{w}(s)) - u(w(\lambda^*, s))\right] dF(s) \\
&\geq \lambda^* \int s\left[u(\tilde{w}(s)) - u(w(\lambda^*, s))\right] dF(s) \\
&\geq \lambda^* \left[\int s\, u(\tilde{w}(s))dF(s) - c'(a^*)\right] \geq 0\,,
\end{aligned}$$

where the second inequality follows because $w(\lambda^*, \cdot)$ satisfies (IC) with equality by assumption, and last inequality follows because $\lambda^* > 0$ by assumption and $\tilde{w}(\cdot)$ satisfies (IC). Notice that $\partial \Pi(w^\epsilon)/\partial \epsilon|_{\epsilon=0} \geq 0$ contradicts (37), so we conclude that $w(\lambda^*, \cdot)$ is uniquely optimal.

Part 2: Strong Duality & Existence

Consider the dual problem

$$\sup_{\lambda\geq 0} L(\lambda, F) = \sup_{\lambda\geq 0} \int \left[w(\lambda, s) + \gamma s^2\right] + \lambda \left[c'(a^*) - su(w(\lambda, s))\right] dF(s)\,, \tag{38}$$

where the equality follows by substituting (8) into the Lagrange function (7). Let λ^* denote the value that maximizes the Lagrange function. Note that $L(\lambda, F)$ is concave in λ, so the first-order condition is necessary and sufficient for an optimal solution. Differentiating $L(\lambda, F)$ with respect to λ, we obtain

$$L_1(\lambda, F) = c'(a^*) - \int su(w(\lambda, s))\, dF(s)\,.$$

Note that $L_1(0, F) = c'(a^*) > 0$, so $\lambda^* > 0$. There are two cases to consider.

First, suppose that there exists some $\lambda^* > 0$, possibly equal to ∞, such that $L_1(\lambda^*, F) = 0$. Then the second term in the brackets of (38) is equal to zero, so $L(\lambda^*, F) = \int \left[w(\lambda^*, s) + \gamma s^2\right] dF(s)$. Notice that $w(\lambda^*, \cdot)$ is feasible, so $\Pi(F) \leq \int \left[w(\lambda^*, s) + \gamma s^2\right] dF(s)$. But weak duality implies that $\Pi(F) \geq L(\lambda^*, F)$, and thus we have $L(\lambda^*, F) = \Pi(F)$. Note that the supremum in (38) need not be attained if $\lambda^* = \infty$.

Second, suppose that $L_1(\lambda, F) > 0$ for all $\lambda \geq 0$. In this case, $\lambda^* = \infty$, but then the second term in the brackets of (38) is strictly positive, so $\sup_{\lambda\geq 0} L(\lambda, F) = \infty$. By weak duality, we have that $\Pi(F) = \infty$ also.

Finally, since strong duality holds, it follows that if $L_1(\lambda^*, F) = 0$, then $w(\lambda^*, s)$ solves (6). If instead $L_1(\lambda, F) > 0$ for all $\lambda \geq 0$, then a solution to (6) does not exist. □

Proof of Lemma 8.

Note that $\underline{s}\,(\lambda)$ and $\overline{s}\,(\lambda)$ are defined by the following equations

$$\begin{aligned} Z_2\left(\lambda, \underline{s}\right) - Z_2\left(\lambda, \overline{s}\right) &= 0, \\ Z\left(\lambda, \underline{s}\right) + \left(\overline{s} - \underline{s}\right) Z_2\left(\lambda, \underline{s}\right) - Z\left(\lambda, \overline{s}\right) &= 0. \end{aligned}$$

The first equation requires that the derivatives of $Z(\lambda, s)$ with respect to s are the same at $s = \underline{s}$ and at $s = \overline{s}$. The second equation requires the point $\left(\overline{s}, Z\left(\lambda, \overline{s}\right)\right)$ lies on the line crossing $\left(\underline{s}, Z\left(\lambda, \underline{s}\right)\right)$ with slope $Z_2\left(\lambda, \underline{s}\right)$. The Jacobian matrix corresponding to this mapping is

$$\begin{vmatrix} Z_{22}\left(\lambda, \underline{s}\right) & -Z_{22}\left(\lambda, \overline{s}\right) \\ -\underline{s} Z_{22}\left(\lambda, \underline{s}\right) & 0 \end{vmatrix}.$$

Since $Z_{22}\left(\lambda, \overline{s}\right) > 0$, the determinant of this matrix is non-zero. Then, by the Implicit Function

Theorem, part (i) of the lemma follows.

To prove part (ii), first, noting that $s_*(\lambda_n)$ converges to $s_*(\lambda^c)$ as $n \to \infty$ and $\underline{s}(\lambda_n) < s_*(\lambda_n) < \overline{s}(\lambda_n)$, it is enough to show that $\overline{s}(\lambda_n) - \underline{s}(\lambda_n)$ tends to zero as $n \to \infty$. Suppose, by contradiction, that there is a subsequence $(\lambda_{n_k})_{n_k} \subset (\lambda_n)_n$ and an $\varepsilon > 0$ such that

$$\overline{s}(\lambda_{n_k}) - \underline{s}(\lambda_{n_k}) > \varepsilon.$$

Therefore, since $s_*(\lambda_{n_k}) \to s_*(\lambda^c)$ and $\underline{s}(\lambda_{n_k}) < s_*(\lambda_n) < \overline{s}(\lambda_{n_k})$, there must exist $s_1, s_2 \in (s_*(\lambda^c) - \varepsilon, s_*(\lambda^c) + \varepsilon)$ and a subsequence $(\lambda_{n_l})_{n_l} \subset (\lambda_{n_k})_{n_k}$ such that $s_2 - s_1 > \varepsilon/2$ and

$$\underline{s}(\lambda_{n_l}) \leq s_1 \text{ and } s_2 \leq \overline{s}(\lambda_{n_l}).$$

Then

$$\begin{aligned} & \lim_{n_l \to \infty} \sup Z_2(\lambda_{n_l}, \underline{s}(\lambda_{n_l})) \leq \lim_{n_l \to \infty} \sup Z_2(\lambda_{n_l}, s_1) = Z_2(\lambda^c, s_1) \\ < \;& Z_2(\lambda^c, s_2) = \lim_{n_l \to \infty} \inf Z_2(\lambda_{n_l}, s_2) \leq \lim_{n_l \to \infty} \inf Z_2(\lambda_{n_l}, \overline{s}(\lambda_{n_l})), \end{aligned}$$

where the first and last inequalities follow from $Z(\lambda_{n_l}, s)$ being convex in s on $(-\infty, \underline{s}(\lambda_{n_l})] \cup [\overline{s}(\lambda_{n_l}), \infty)$, the two equalities follow from continuity, and the strict inequality follows from $Z(\lambda^c, s)$ being strictly convex (see Lemma 6). Note, however, that $Z_2(\lambda_{n_l}, \underline{s}(\lambda_{n_l})) = Z_2(\lambda_{n_l}, \overline{s}(\lambda_{n_l}))$ and hence

$$\lim_{n_l \to \infty} \sup Z_2(\lambda_{n_l}, \underline{s}(\lambda_{n_l})) \geq \lim_{n_l \to \infty} \inf Z_2(\lambda_{n_l}, \overline{s}(\lambda_{n_l})),$$

which contradicts the previous displayed inequality chain. □

Proof of Lemma 9.

First note that if the IC is *not* slack for a given λ then the Principal's payoff is bounded from below by $\underline{w}$. Indeed, even if the Principal does not acquire any information, he has to pay at least $\underline{w}$ to the Agent. Therefore, in order to prove the lemma, it is enough to show that if λ is large enough then the Principal's payoff is smaller than $\underline{w}$.

Let $\overline{u}$ denote $\lim_{w \to \infty} u(w)$ and fix an $\widetilde{s} > 0$ such that

$$\widetilde{s} > \frac{2c'(a^*)}{\overline{u} - u(\underline{w})}. \tag{39}$$

(If $\overline{u} = \infty$ then this inequality imposes no restriction on $\widetilde{s}$ in addition to requiring it to be positive.) Consider the binary distribution, $\widetilde{F}$, which specifies probability half on $\widetilde{s}$ and $-\widetilde{s}$.[17] Recall that

$$\frac{\partial \mathbb{E}_{\widetilde{F}}\left[Z(\lambda,s)\right]}{\partial\lambda} = -\int su\left(w\left(\lambda,s\right)\right)d\widetilde{F}\left(s\right)+c'(a^*)$$
$$= \frac{1}{2}\widetilde{s}u\left(\underline{w}\right)-\widetilde{s}\frac{1}{2}u\left(w\left(\lambda,\widetilde{s}\right)\right)+c'(a^*).$$

Since $\lim_{\lambda\to\infty} w\left(\lambda,\widetilde{s}\right)=\infty$, $\lim_{\lambda\to\infty} u\left(w\left(\lambda,\widetilde{s}\right)\right)=\overline{u}$. Therefore,

$$\lim_{\lambda\to\infty}\frac{\partial \mathbb{E}_{\widetilde{F}}\left[Z(\lambda,s)\right]}{\partial\lambda} = -\frac{\widetilde{s}\left[\overline{u}-u\left(\underline{w}\right)\right]}{2}+c'(a^*)<0,$$

where the inequality follows from (39). (If $\overline{u}=\infty$ then this limit is minus infinity.) Since $w\left(\lambda,\widetilde{s}\right)$ is strictly increasing in λ it follows that there exists a $\overline{\lambda}$ such that for all $\lambda>\overline{\lambda}$,

$$\frac{\partial \mathbb{E}_{\widetilde{F}}\left[Z(\lambda,s)\right]}{\partial\lambda}<0.$$

Since $\partial\mathbb{E}_{\widetilde{F}}\left[Z(\lambda,s)\right]/\partial\lambda$ is strictly decreasing in λ (because $u\left(w\left(\lambda,\widetilde{s}\right)\right)$ is strictly increasing in λ), it follows that there exists a Λ such that the Principal's payoff is smaller than $\underline{w}$ whenever $\lambda>\Lambda$ and the Principal chooses $\widetilde{F}$. Of course, the Principal's payoff is even smaller if she best-responds to λ. $\square$

[17] $\widetilde{F}(s)=\begin{cases}0 & \text{if } s<-\widetilde{s}\\ \frac{1}{2} & \text{if } s\in[-\widetilde{s},\widetilde{s})\\ 1 & \text{if } s\geq\widetilde{s}.\end{cases}$