fedea

Fundación de Estudios de Economía Aplicada

Executive Pay with Observable Decisions

by

Marco Celentani[*]

Rosa Loveira[**]

Pablo Ruiz-Verdú[***]

Documento de Trabajo 2010-29

Pymes

CÁTEDRA Fedea-Banesto

December 2010

[*] Department of Economics, Universidad Carlos III and Fedea.

[**] Universidade de Vigo.

[***] Universidad Carlos III.

Los Documentos de Trabajo se distribuyen gratuitamente a las Universidades e Instituciones de Investigación que lo solicitan. No obstante están disponibles en texto completo a través de Internet: *http://www.fedea.es.*

These Working Paper are distributed free of charge to University Department and other Research Centres. They are also available through Internet: *http://www.fedea.es.*

ISSN:1696-750X

Jorge Juan, 46

28001 Madrid -España

Tel.: +34 914 359 020

Fax: +34 915 779 575

infpub@fedea.es

N.I.F. G-78044393

EXECUTIVE PAY WITH OBSERVABLE DECISIONS *

Marco Celentani[†] Rosa Loveira[‡] Pablo Ruiz-Verdú[§]

October 2010

ABSTRACT: We propose a model of delegated expertise designed to analyze executive compensation. An expert has to pick one of two possible decisions. By exerting effort the expert can obtain private information on these decisions. The expert's decision and its ultimate performance realization are publicly observed, but the expert's information is not. In other words, the principal observes the expert's decision and its realization, but does not know whether the expert expended effort to obtain information and whether he made an efficient decision conditional on the information he received. We characterize the optimal compensation contract among those that give the expert incentives to obtain information to determine the efficient decision and to make the decision that is efficient contingent on the obtained information. We show that: 1) It is generically optimal to make pay contingent on the decision made by the expert, not only on performance; 2) The expert is often rewarded for choosing alternatives that are ex-ante inefficient. 3) When decisions differ in their complexity, optimal pay-performance may be zero if the expert chooses the complex alternative. Our model highlights novel factors that should be considered in the design of executive compensation contracts, sheds light on existing compensation practices, such as rewarding executives for acquisitions, and suggests mechanisms to promote managerial innovation.

KEYWORDS: Delegated expertise, executive compensation.

JEL CLASSIFICATION NUMBERS: D86, M12, M52

*We gratefully acknowledge the financial support of FEDEA and the Spanish Ministry of Innovation and Science for financial support under grant SEJ2008-03516. Rosa Loveira gratefully acknowledges the financial support of Xunta de Galicia (Spain) under the Isidro Parga Pondal research grant. We thank seminar participants at Universidad Carlos III, Universidade de Vigo, Econometric Society European Meeting 2007, 2nd European Reward Management Conference 2009, EASSET 2009, EARIE 2010 and Jornadas de Economía Industrial 2010 for useful discussions and suggestions.

[†]Department of Economics, Universidad Carlos III; marco.celentani@uc3m.es

[‡]Departamento de Fundamentos del Análisis Económico e Historia e Instituciones Económicas, Universidade de Vigo; rloveira@uvigo.es.

[§]Department of Business Administration, Universidad Carlos III; pablo.ruiz@uc3m.es.

1 Introduction

By the very nature of the CEO's role, most of the key decisions that a CEO makes can be observed by shareholders. Thus, shareholders can observe the markets a firm enters or exits, the firm's choice of capital structure, or whether a firm acquires another firm or spins off a division. Therefore, CEO pay could be made to depend on observable decisions made by the CEO, and not only on the outcome, in terms of the firm's financial performance, of those decisions. In fact, CEOs are often rewarded for the decisions they make. For example, CEOs are awarded bonuses for acquiring other firms (Grinstein and Hribar, 2004; Harford and Li, 2007), or for initiating joint ventures, spin-offs, strategic alliances, or seasoned equity offerings (Fich et al., 2008). However, the problem with conditioning managers' pay on their decisions is that, often, shareholders do not know which decisions are optimal. Indeed, as Murphy (1999) puts it, "the reason shareholders entrust their money to self-interested CEOs is based on shareholder beliefs that CEOs have superior skills or information in making investment decisions".[1] Thus, conditioning pay on decisions runs the risk of providing the wrong incentives to the manager. As a case in point, the practice of rewarding CEOs for acquisitions has been criticized (Bebchuk and Fried, 2004; Grinstein and Hribar, 2004) on the grounds that acquisitions do not, on average, benefit the shareholders of the acquiring firm (Andrade et al., 2001; Moeller et al., 2004 and 2005).

In this paper, we analyze whether, in a context of uncertainty about the optimal decision, managers should be rewarded for the decisions they make or exclusively for the returns that those decisions generate for shareholders. We find that CEOs' pay should generally be conditioned on their observable decisions, and we characterize the relation between firm performance, managerial decisions, and pay prescribed by the optimal compensation contract.

The usual analytical framework to analyze executive compensation is the principal-agent model with hidden action. In this model, the principal knows the action that she would like the agent to take, but cannot make compensation contingent on the action taken by the agent because this action is not observable. The problem for the principal is to design a compensation contract that, by conditioning pay on observables—which typically include some measure of firm performance—provides incentives to the manager to take the right action at the minimum cost for the principal. Certain aspects of the general problem of providing incentives to executives fit this framework well. For example, executives may take actions that are difficult to observe and that

[1] Murphy (1999), page 2521.

reduce firm value, like diverting firm funds for his personal use, or enjoying excessive perquisites (Jensen and Meckling, 1976). In this case, the action that maximizes firm value is known in advance (no fund diversion, no superfluous perquisites), and the problem for shareholders is to design the compensation contract in such a way that the manager refrains from taking (unobservable) inefficient actions.

The standard hidden action framework is, however, only of limited value to analyze the problem of providing incentives to CEOs to ensure that they make the right strategic decisions. These strategic decisions, such as determining whether to acquire a supplier or merge with a rival, determine firm value to a much larger extent than the manager's enjoyment of excessive private benefits. Strategic decisions are, however, typically observable, so that—unlike in the hidden action model—the problem for shareholders is not how to convince the manager to implement a decision that is known to be efficient, but, rather, how to provide incentives for the manager to learn which decision is efficient, and, then, to effectively make the efficient decision.[2] The main agency problem between shareholders and management, thus, arguably arises not because shareholders do not observe managerial actions but because shareholders do not know which actions managers should take in order to maximize firm value. While the standard hidden-action model is well suited to agency problems in which the principal knows what is best, but cannot observe what the agent carries out the task, it is not so well suited to study the moral hazard problem of top executives.

In this paper, we analyze the relationship between a risk-neutral principal (shareholders) who can observe the decisions made by a risk-averse expert agent (the CEO), but does not know which decision is efficient nor whether the expert exerted the necessary effort to learn the efficient decision. The problem for the principal is how to design a compensation contract to provide incentives for the expert to exert effort to obtain information and to make an efficient decision based on that information. We study this problem in a simple framework in which there are only two alternative decisions and two possible outcomes (success or failure) of those decisions.

Given that the expert's decisions are contractible, the first question we address is whether these observable decisions should enter the compensation contract at all or whether pay should depend only on the firm's profits. The usual guide to determine the variables to be included in a

[2]This point has also been made by Murphy (1999): "Unobservable actions cannot be the driving force underlying executive contracts: even if shareholders (or boards of directors) could directly monitor CEO actions, they could not tell whether the actions were appropriate given the circumstances." (Murphy, 1999, page 2521.)

contract is the informativeness—or sufficient statistic—principle (Holmstrom, 1979). According to this principle, a variable should be included in an executive compensation contract if it provides information as to whether the executive took the desired action. The informativeness principle, however, cannot be applied in our context to determine whether the variable of interest—the CEOs' decision— should enter the contract. The reason is that this variable is itself chosen by the CEO. Therefore, even if certain decisions may signal that the CEO put effort into information acquisition, rewarding those decisions may distort the CEO's incentives to make the efficient decision, and these incentives are key in a context in which shareholders do not know the efficient decision.

Our model yields several novel results about the design of optimal compensation contracts. First, we show that it is generically optimal to make the expert's pay depend on observable decisions even if shareholders do not know which is the efficient decision. Even if the expert could be given the right incentives by means of a contract that made his pay contingent on profits only, making pay depend on observable decisions allows the firm to provide those incentives at a lower cost by reducing the expert's risk exposure.

Second, we show that the form of the optimal compensation contract crucially depends on two factors. The first factor is the precision with which the informed expert can predict the outcome of each decision. For some decisions the information acquired by the expert may allow him to form precise predictions of the decision's outcome, while for other decisions the information acquired by the expert may be only a noisy signal of the decision's outcome. Thus, managerial learning may be symmetric (the informed expert can predict the outcome of each decision with similar precision) or asymmetric (the informed expert can predict the outcome of a decision with much greater precision than the outcome of the other). Learning asymmetry may arise because some decisions are more complex or there is less prior knowledge about them that the expert can draw upon to interpret the information he receives. Alternatively, the expert may be specialized in assessing the consequences of one of the available decisions, but may not be able to obtain a significant informational advantage with respect to shareholders regarding the other decision. For example, a manager may be a takeover specialist, who is able to determine with great precision the outcome of the acquisition of a supplier firm, yet may not be better equipped than board directors to assess the outcome that would result if the firm did not carry out the acquisition. We show that the form of the optimal contract depends on the symmetry of managerial learning. In particular, when learning is sufficiently asymmetric, compensation is completely insensitive to performance if the manager makes the decision whose outcome is more uncertain. Thus, in some cases, observed pay-performance sensitivity may be low even if the manager's contract is set optimally. In these cases, the incentives to make the right decision are provided by the performance sensitivity of pay had the alternative decision been made.

The second factor that determines the form of the compensation contract is shareholders' (or board directors) ex ante information about the optimal decision. Even if board directors do not know which decision is efficient, they need not think that all decisions are equally likely to be efficient. For example, on the basis of existing empirical research on the effect of acquisitions on the stock price of the acquiring firm, directors may expect that acquisitions are likely to be suboptimal, even if they understand that under certain circumstances it may be optimal to undertake an acquisition. We show that the ex ante information about the efficient decision determines the form of the compensation contract. In particular, when the precision with which the manager learns about the two decisions is similar, the manager is optimally paid more (conditional on performance) if he takes the decision that is considered to be ex ante inefficient. Therefore, our results show that it may be optimal to reward managers for the decisions they make. Moreover, it may be optimal to reward them for making decisions that, on average, can be expected to be wrong. It follows from these results that it may be optimal to reward managers for acquiring other firms, even though, on average, acquisitions have been shown not to increase—and to greatly reduce in the case of large operations (Moeller et al., 2005)—the wealth of the acquiring firm's shareholders. However, we stress that this is a normative point and we do not argue that the observed pay premium for acquisitions is the outcome of an optimal contract.

The paper is organized as follows. Section 2 presents the basic model for the case in which the realizations of the possible decisions made by the expert are independent. In Section 3 we prove that generically there is a gain in making pay conditional not only on performance but also on the expert's observable decisions. In Sections 4 and 5 we characterize the way in which pay should depend on observable decisions and performance. In Section 6 we consider the case in which the the realizations of the possible decisions made by the expert are negatively correlated. In Section 7 we consider the case of uninformative priors, i.e., the case in which the two possible decisions are ex-ante equivalent. Section 8 discusses the implications of our results and applies them to shed light on how managers' pay should depend on acquisition and layoff decisions. Finally, Section 9 summarizes the results and presents some concluding remarks. All proofs are presented in the Appendices.

2 The Model

2.1 Sequence of events

A risk neutral principal has to choose one of two common knowledge independent probability distributions from the set $\{A, B\}$. Each of the two probability distributions has support on $\{S, F\}$, with $\{S, F\} \in \mathbb{R}^2$ and $F < S$. The two distributions may represent the consequences of entering one of two markets, of acquiring a potential target or not, or of performing one of two treatments on a patient. We denote by r_A and r_B the realizations of probability distribution A and B, respectively. We denote by a the probability of success under A, and by b the probability of success under B. The principal knows the two probability distributions and therefore has expectations on which of them is efficient. But she can also hire an expert agent to assist her with the decision. In the rest of the paper we will refer to the latter as simply the expert. If the principal hires the expert, the latter makes a decision to shirk ($e = 0$) or expend effort ($e = 1$) that is unobservable to the principal. If he decides to shirk, he obtains no additional information about A and B. But if he expends effort, he obtains two independent signals on A and B, respectively α and β. We assume that $\alpha \in \{\overline{\alpha}, \underline{\alpha}\}$ and $\beta \in \left\{\overline{\beta}, \underline{\beta}\right\}$, where $\overline{\alpha}$ is a favorable signal on A, $\underline{\alpha}$ is an unfavorable signal on A, and similarly for β and B. If hired, the expert also chooses the probability distribution. If he did expend effort and obtained the two signals, he can make the choice of the probability distribution contingent on the signals he has observed. We assume that the expert's choice of a probability distribution is observable and contractible and in the rest of the paper we will refer to the choice of a probability distribution as a *decision.*

Note that in line with the literature on delegated expertise we assume that (i) expending effort provides the expert with a superior ability to forecast the realizations of the probability distributions, but does not affect the probability distributions themselves; (ii) communication between the principal and the expert is too costly, so that, if the principal hires the expert, she also delegates the choice of the probability distribution to him.

We assume that the expert has reservation utility $\overline{U}$ and has preferences described by the following Bernoulli utility function $U(w) - g(e)$, with $g(1) - g(0) = g > 0$ and $U(.)$ is continuously differentiable, with $U'(.) > 0$ and $U''(w) < 0$. In other words, we assume that the expert is strictly risk averse.

At the contracting stage, we assume that the principal makes a take-it-or-leave-it contract offer to the expert. A contract specifies a salary payment for every possible public history following acceptance of the contract.

The extensive form of the game is summarized as follows:

1. The principal makes a take-it-or-leave-it contract offer $w = (w_{AF}, w_{AS}, w_{BF}, w_{BS}) \in \mathbb{R}^4$. Each of the elements of w represents the salary payment for each of the possible public histories following the acceptance of the contract by the expert (combinations of a decision and a realization of return).

2. The expert accepts or rejects.

   (a) If the expert rejects, the game ends and he obtains reservation utility $\overline{U}$.

   (b) If the expert accepts, he is hired.

      i. The expert chooses whether to exert effort, $e = 1$, or not, $e = 0$.

      ii. Nature privately chooses the vector of return realizations $r = (r_A, r_B) \in \{F, S\}^2$.

      iii. If the expert has not exerted effort (information set 0), he makes a decision $d \in \{A, B\}$.

      iv. If the expert has exerted effort, he receives signals α and β. We denote a generic realizations of signals

      v. The expert makes a decision $d \in \{A, B\}$.

      vi. The expert's decision and the return realization are publicly observed. The public history following acceptance of the contract by the expert is $(d, r_d) \in \{A, B\} \times \{S, F\}$.

      vii. The principal pays the expert the salary associated with the realized public history.

Without loss of generality we assume that $a \geq b$, but we will often refer to the generic case in which $a > b$.

Assumption 1 *Decision A is ex ante efficient: $a > b$.*

2.2 Signals

If the expert exerts effort, he receives signals α and β. We assume that the distributions of α and β conditional on the realizations of A and B are

$$\Pr\left(\alpha = \overline{\alpha} \mid (r_A, r_B)\right) = \begin{cases} 1-\varepsilon_A & \text{if } r_A = S \\ \varepsilon_A & \text{if } r_A = F \end{cases} ; \Pr\left(\beta = \overline{\beta} \mid (r_A, r_B)\right) = \begin{cases} 1-\varepsilon_B & \text{if } r_B = S \\ \varepsilon_B & \text{if } r_B = F \end{cases} . \tag{1}$$

$\varepsilon_d \in \left[0, \frac{1}{2}\right]$, $d \in \{A, B\}$, measures the precision of the expert's information regarding the outcome of decision d. If $\varepsilon_d = 0$, the expert can perfectly forecast the outcome of decision d; if $\varepsilon_d = 1/2$, the signal about decision d is completely uninformative. Notice that we assume that the signals are independently distributed conditional on the realizations of returns.

We denote by $\overline{a}$ the probability of S under A conditional on the favorable signal, $\overline{\alpha}$

$$\overline{a} = \Pr\left(r_A = S|\overline{\alpha}\right) = \frac{\Pr\left(\overline{\alpha}|r_A = S\right)\Pr\left(r_A = S\right)}{\Pr\left(\overline{\alpha}|r_A = S\right)\Pr\left(r_A = S\right) + \Pr\left(\overline{\alpha}|r_A = F\right)\Pr\left(r_A = F\right)}$$

In a similar way we define $\underline{a}$, $\overline{b}$, and $\underline{b}$. Notice that $\varepsilon_d \in \left[0, \frac{1}{2}\right]$, $d \in \{A, B\}$ implies that $\overline{a} > \underline{a}$ and $\overline{b} > \underline{b}$, so that $\overline{\alpha}$ and $\overline{\beta}$ should be interpreted as favorable signals on A and B, respectively, and $\underline{\alpha}$ and $\underline{\beta}$ should be interpreted as unfavorable signals.

We denote by $\hat{\delta}(\alpha, \beta)$ the set of efficient decisions given $(\alpha, \beta) \in \{\overline{\alpha}, \underline{\alpha}\} \times \left\{\overline{\beta}, \underline{\beta}\right\}$. Notice that $d \in \hat{\delta}(\alpha, \beta)$ is equivalent to

$$\Pr\left(r_d = S|\left(\alpha, \beta\right)\right) \geq \Pr\left(r_{d'} = S|\left(\alpha, \beta\right)\right), d \neq d'.$$

Lemma 1 *Under Assumption 1,* $\hat{\delta}\left(\overline{\alpha}, \underline{\beta}\right) = A$.

For given $\varepsilon_B \in \left[0, \frac{1}{2}\right]$, we define

$$\underline{\varepsilon}_A\left(\varepsilon_B\right) \equiv \frac{(1-a)\, b\varepsilon_B}{(1-a)\, b\varepsilon_B + a\,(1-b)\,(1-\varepsilon_B)} \tag{2}$$

and

$$\overline{\varepsilon}_A\left(\varepsilon_B\right) \equiv \frac{a\,(1-b)\,\varepsilon_B}{a\,(1-b)\,\varepsilon_B + (1-a)\, b\,(1-\varepsilon_B)}. \tag{3}$$

Notice that under Assumption 1 $0 < \underline{\varepsilon}_A\left(\varepsilon_B\right) < \varepsilon_B < \overline{\varepsilon}_A\left(\varepsilon_B\right)$, but notice also that $\overline{\varepsilon}_A\left(\varepsilon_B\right)$ may be larger than $\frac{1}{2}$.

Assumption 2 $\varepsilon_A < 1 - \overline{\varepsilon}_A\left(\varepsilon_B\right)$.

Lemma 2 *Under Assumption 2,* $\hat{\delta}\left(\underline{\alpha}, \overline{\beta}\right) = B$.

Lemma 2 means that with Assumption 2 we assume that signals are sufficiently informative, so that an unfavorable signal about the ex ante efficient decision A and a favorable signal about the ex ante inefficient decision B imply that decision B is efficient. This is important, because, otherwise, decision A would be efficient irrespectively of the signal received by the expert and it would not make sense to consider how to induce the expert to exert effort to obtain signals.

DEFINITION 1 *1. The efficient decision is* sensitive to α *if* $\hat{\delta}(\alpha, \beta)$ *is constant in* β. *2. The efficient decision is* sensitive to β *if* $\hat{\delta}(\alpha, \beta)$ *is constant in* α. *3. The efficient decision is* sensitive to α and β *if* $\hat{\delta}(\alpha, \beta)$ *is constant in neither* α *nor* β.

The following lemma characterizes which decisions are efficient when both signals are either favorable or unfavorable, depending on the values of ε_A and ε_B. For simplicity and without loss of generality,, when $\varepsilon_A = \underline{\varepsilon}_A(\varepsilon_B)$ or $\varepsilon_A = \overline{\varepsilon}_A(\varepsilon_B)))$ we assume $\hat{\delta}\left(\overline{\alpha}, \overline{\beta}\right) = A$ and $\hat{d}\left(\underline{\alpha}, \underline{\beta}\right) = A$ as an indifference-breaking rule.

LEMMA 3 *Under Assumptions 1 and 2:*

1. *The efficient decision is sensitive to* α *if and only if* $\varepsilon_A < \underline{\varepsilon}_A(\varepsilon_B)$; *in this case* $\hat{d}\left(\overline{\alpha}, \overline{\beta}\right) = A$ *and* $\hat{d}\left(\underline{\alpha}, \underline{\beta}\right) = B$.

2. *The efficient decision is sensitive to* β *if and only if* $\varepsilon_A > \overline{\varepsilon}_A(\varepsilon_B)$; *in this case* $\hat{d}\left(\overline{\alpha}, \overline{\beta}\right) = B$ *and* $\hat{d}\left(\underline{\alpha}, \underline{\beta}\right) = A$.

3. *The efficient decision is sensitive to* α *and* β *if and only if* $\varepsilon_A \in [\underline{\varepsilon}_A(\varepsilon_B), \overline{\varepsilon}_A(\varepsilon_B)]$; *in this case* $\hat{d}\left(\overline{\alpha}, \overline{\beta}\right) = A$ *and* $\hat{d}\left(\underline{\alpha}, \underline{\beta}\right) = A$.

Lemma 3 means that when the precision of α (the signal on decision A) is high enough relative to the precision of β ($\varepsilon_A < \underline{\varepsilon}_A(\varepsilon_B)$) observing β is irrelevant for determining the efficient decision: If $\alpha = \overline{\alpha}$, the efficient decision is A, and if $\alpha = \underline{\alpha}$, the efficient decision is B. Similarly, if the precision of α is low enough relative to the precision of β ($\varepsilon_A > \overline{\varepsilon}_A(\varepsilon_B)$), observing β is sufficient to determine the Efficient decision. In either of these two cases, the expert could optimally discard one of the two signals. However, when the two signals have a similar precision ($\underline{\varepsilon}_A(\varepsilon_B) < \varepsilon_A < \overline{\varepsilon}_A(\varepsilon_B)$), observation of both signals is necessary to determine the efficient decision. In particular, when the two signals have similar precision, observing $(\alpha, \beta) = \left(\overline{\alpha}, \overline{\beta}\right)$ or $(\alpha, \beta) = \left(\underline{\alpha}, \underline{\beta}\right)$ does not lead the expert to revise the a priori ordering of the decisions. Thus, knowing that $\alpha = \underline{\alpha}$ (or that $\beta = \overline{\beta}$) is not sufficient to infer that the efficient decision is B.

Notice also that under Assumptions 1 and 2 and the indifference-breaking rule $\hat{d}(\alpha,\beta)$ is single valued for every $(\alpha,\beta) \in \{\overline{\alpha},\underline{\alpha}\} \times \{\overline{\beta},\underline{\beta}\}$.

2.3 The optimal contracting problem

We define an optimal contract as a contract that minimizes the expected salary payment to the expert among all contracts thar are (i) accepted by the expert; (ii) such that the expert weakly prefers exerting effort to shirking; and (iii) such that the expert weakly prefers to make efficient decisions conditional on the private information he obtains.

Let $\delta_0 \in \{A,B\}$ be the expert's action following acceptance and shirking. In words, d_0 describes the expert's action when he shirks and therefore obtains no private signal. Let $\delta : \mathcal{A} \times \mathcal{B} \longrightarrow \{A,B\}$ be the expert's action following acceptance and exerting effort.[3] In words, $\delta(\alpha,\beta)$ describes the expert's action conditional on (α,β), the private signal obtained by exerting effort. The expert's action profile following acceptance of a contract can therefore be denoted as $s = (e,\delta_0,\delta)$. In the following we focus on contracts that once accepted, induce the expert to play $s = (e,\delta_0,\delta) = \hat{s} = \left(1,A,\hat{d}\right)$. In words, we focus on contracts that are such that the expert exerts effort ($e=1$), obtains the private signal and makes the efficient decision conditional on the private signals ($\delta = \hat{\delta}$). Notice that we also require the efficient decision to be made conditional on exerting no effort, ($\delta = A$) (under Assumption 1 A is ex-ante efficient). But this requirement is unnecessary and has no consequences. The reason is that if the expert has incentives to exert efforts, there is 0 probability that he will not exert it and that he will need to make a decision based only on prior information.

Denote by $E(w|s)$ the expected wage payment conditional on the expert accepting the contract and playing s. Denote by $V_0(w,\delta)$ the expert's ex ante expected utility of salary payments after having accepted contract w, after having exerted effort, and conditional on playing δ. Denote by $V_1(w,d|(\alpha,\beta))$ the expert's expected utility from decision d conditional on having observed signal (α,β). An optimal contract is thus the one that solves the following problem:

[3] We can restrict ourselves to pure strategies, since we will focus on contracts that induce effort and implement the Efficient decision given the information obtained by the expert. Generically, this Efficient decision will be unique, so that randomization does not take place in equilibrium. For the incentive compatibility constraints, we need only care about deviations to other pure behavior strategies.

$$
\begin{aligned}
\min_{w\in\mathbb{R}^4} \quad & E\left(w|\hat{s}\right) && \text{(P)}\\
\text{s.t.} \quad & V_0\left(w,\hat{s}\right) - g\left(1\right) \geq \overline{U} && \text{(PC)}\\
& V_0\left(w,\hat{s}\right) - g\left(1\right) \geq aU\left(w_{AS}\right) + (1-a)U\left(w_{AF}\right) - g\left(0\right) && \text{(IC-}A\text{)}\\
& V_0\left(w,\hat{s}\right) - g\left(1\right) \geq bU\left(w_{BS}\right) + (1-b)U\left(w_{BS}\right) - g\left(0\right) && \text{(IC-}B\text{)}\\
& V_1\left(w,\hat{\delta}\left(\alpha,\beta\right)|\left(\alpha,\beta\right)\right) \geq V_1\left(w,d|\left(\alpha,\beta\right)\right),\ d\neq\hat{\delta}\left(\alpha,\beta\right),\ (\alpha,\beta)\in\mathcal{A}\times\mathcal{B} && \text{(IC-}\alpha\beta\text{)}
\end{aligned}
$$

Constraint (PC) is a participation constraint; (IC-A) and (IC-B) are two incentive compatibility constraints that require that the expert weakly prefers to play $\hat{s}$ over exerting no effort and making either of the decisions. Each of (IC-$\alpha\beta$), $(\alpha,\beta)\in\mathcal{A}\times\mathcal{B}$, is an incentive compatibility constraint that requires that, conditional on each of the possible signal he may receive, the expert weakly prefers to make the efficient decision $\hat{\delta}\left(\alpha,\beta\right)$.

The assumption on the expert's preferences (U is continuously differentiable, $U'>0$ and $U''<0$) guarantee that a solution to the previous problem exists and is unique[4]. For this reason in the rest of the paper we will simply be concerned with the characterization of the optimal contract, i.e., the unique solution to the previous problem.

3 Conditioning pay on decisions

The traditional guide to determine which performance measures to include in a compensation contract is Holmström's (1979) informativeness principle. The informativeness principle holds that a performance measure should be included in the contract whenever it provides information on the action taken by an agent beyond the information provided by other measures already in the contract. As discussed in the introduction, however, there are important limitations to the use of the informativeness principle, since it is derived in a context in which the principal knows the action that she would like the agent to take. Further, the additional measures are supposed to be stochastically related to the agent's action and not themselves chosen by the agent. Neither of these assumptions holds in our model, since the principal does not know the efficient decision and since the additional measure considered is itself a decision made by the expert agent.

[4]To see this notice first that one can rewrite the previous problem in terms of expert utilities. By $U''<0$ the objective function would be strictly quasi-convex and the constraints would be linear and therefore continuous and convex.

Before characterizing the optimal contract, we first address the question as to whether the expert's pay should be made contingent on the expert's decisions or only on outcomes:

PROPOSITION 1 *Let* $w^* = (w^*_{AF}, w^*_{AS}, w^*_{BF}, w^*_{BS})$ *be the optimal contract. If* $a \neq b$*, then, either* $w^*_{AF} \neq w^*_{BF}$ *or* $w^*_{AS} \neq w^*_{BS}$ *(or both). If* $a = b$ *and* $\varepsilon_A = \varepsilon_B$*, then* $w_{AS} = w_{BS} > w_{AF} = w_{BF}$*.*

Proposition 1 states that as long as one of the two decisions has a higher ex ante probability of success, it is optimal to condition the expert's pay not only on profits, but also on the decision made by the expert. Only in the case in which both decisions are considered ex ante to have exactly the same probability of success it may be optimal to condition the expert's pay only on returns, a possibility that we discuss further in Section 7. As mentioned in the introduction, Murphy (1999) has argued, in his oft-cited review of the executive compensation literature, that uncertainty about the efficient decision would lead to contracts that are based on the principal's objective. Proposition 1 shows that this is the case only if there is extreme uncertainty about the efficient decision, i.e., only if the principal's prior assigns exactly the same probability of success to both decisions.

To understand Proposition 1, it is worth noting that any contract w that conditions pay only on returns ($w_{AF} = w_{BF} = w_F < w_S = w_{AS} = w_{BS}$) will give the expert the incentives to make the efficient decision contingent on the signal received, since the efficient decision by definition leads to a higher probability of success. It is possible, however, to maintain those incentives while reducing the risk faced by the expert by conditioning pay on the expert's decision. Since the expert is risk averse, this reduces the cost of the contract to the principal. To see this, take any contract w that satisfies the participation constraint and all incentive compatibility constraints, and that does not condition pay on the expert's decision. This contract must establish $w_S > w_F$, so that all second-stage incentive compatibility constraints hold with slack. Moreover, since $a > b$, ensuring that the expert prefers to exert effort rather than shirking and choosing A also ensures that he prefers to exert effort rather than shirking and choosing B. Starting from this contract, it is possible to compress the expert's pay while ensuring that the participation constraint and all incentive compatibility constraints still hold. A possible way to do this is to reduce pay in case of success if the expert takes decision A and to increase pay in case of failure if the expert takes decision B in such a way that the ex ante expected utility of the expert is not changed. Since the expert is risk averse, this reduces the expected cost of the contract. Further, for small enough changes, all but one of the incentive compatibility constraints will still hold, since they held with slack for contract w. The only incentive compatibility constraint that could possibly be binding—the one ensuring that the expert prefers to exert effort over exerting no effort and choosing A—will also hold, since the proposed deviation does not change the expert's expected utility, yet it reduces the expected utility of exerting no effort and choosing A. It is worth noting that, even though the proposed deviation would implement the efficient action profile at a lower cost than a contract that does not condition pay on the expert's decision, it does not follow that the optimal contract necessarily entails $w_{AS} < w_{BS}$ and $w_{AF} < w_{BF}$, as implied by the proposed deviation.

4 Rewarding decisions

Consider the benchmark case in which both signals are equally informative ($\varepsilon_A = \varepsilon_B = \varepsilon$). In this case, it follows from Lemma 3 that if both signals have the same sign, the ex-post efficient decision is the ex ante efficient decision A. Thus, decision B is efficient only conditional on $(\underline{\alpha}, \overline{\beta})$.

Proposition 2 *The optimal contract w^* satisfies:*

$$w^*_{AF} < w^*_{BF} < w^*_{AS} < w^*_{BS}.$$

Therefore, pay is increasing in returns, both conditionally on the decision made ($w^*(d, F) < w^*(d, S)$, for any d) and unconditionally ($w^*(d, F) < w^*(d', S)$, for any $d \neq d'$). This, by itself, implies that, if the expert exerts effort, he has the incentive to make the decision that is efficient given the signal.

In line with by Proposition 1, the expert's pay depends on the decision made ($w^*(A, r) \neq w^*(B, r)$, for any $r \in \{F, S\}$). Moreover, pay is monotonic in the decision made, in the sense that, conditional on a given outcome, one decision always leads to strictly higher pay. Further, the decision that leads to greater pay is ex ante inefficient. Thus, when A is the ex ante efficient decision, the optimal contract pays the expert more, conditional on returns, if he takes decision B. This is noteworthy, as it implies that the expert is paid more for making decisions that are wrong on average (and that are known to be wrong on average by the principal). The rationale for the result, however, is compelling. If pay is weakly increasing in performance, then an uninformed expert will always find it optimal to make the ex ante efficient decision. Only an informed expert could possibly choose the ex ante inefficient decision if, upon exerting effort, he learnt that it could be expected to be superior to the alternative. To encourage the expert to exert effort, it is, therefore, optimal for the principal to increase the payoff of taking the ex ante inefficient decision, since, by doing so, the principal increases the expected utility of an expert that exerts effort while keeping the expected utility of an uninformed expert constant.

An alternative way to understand Proposition 2 is to consider first the contract that would be optimal if the expert always made the efficient decision contingent on his private information, so that the only problem for the principal would be to induce effort provision. In this case, an expert who shirks will never play B, the ex-ante inefficient decision. Therefore, the likelihood of generating public history BF or BS (i.e., a public history in which decision B takes place) relative to any other public history is higher when the expert exerts effort than when he shirks (in fact, it is infinitely higher). This implies that, in order to induce effort, the expert's pay after these public histories should be higher than after any other public history. Notice also that, given that an expert who shirks never generates public histories public history BF or BS, a contract with $w_{BF} \neq w_{BS}$ will be dominated by another contract with $w_{BF} = w_{BS}$, since the latter reduces the risk faced by an expert who exerts effort and, thus, reduces the cost of the contract. Finally, note that the likelihood of history AS relative to history AF is higher when the expert exerts effort, so the contract should specify $w_{AS} > w_{AF}$. The optimal ranking of salaries would, thus, be:

$$w_{AF} < w_{AS} < w_{BF} = w_{BS} \tag{4}$$

With this ranking, however, the expert would always choose B. In order to ensure that the expert chooses A when he receives signal $(\overline{\alpha}, \underline{\beta})$, it is necessary to make $w_{AS} > w_{BF}$. Thus, we obtain the ranking derived in Proposition 2:

$$w_{AF} < w_{BF} < w_{AS} < w_{BS} \tag{5}$$

5 Decision uncertainty/complexity and pay-performance sensitivity

An expert is often able to predict the outcomes of different decisions with different degrees of precision. For example, consider a firm that has to decide whether to enter a market that is similar to its home market or a completely new market. Even if the manager allocates his effort optimally to the tasks of investigating the consequences of each alternative, it is likely that in the end he will have a more precise profit forecast for the known market than for the completely new market. Thus, an asymmetry in the manager's ability to learn the consequences of the different alternatives may arise because some decisions are more complex or there is less prior knowledge about them that the manager can draw upon to interpret the information he receives. But the asymmetries in the precision with which the manager can forecast the outcomes of different decisions may also be idiosyncratic to the manager. A manager may be specialized in assessing the consequences of one of the available decisions, but may not be able to obtain a significant informational advantage with respect to shareholders regarding an alternative decision. Continuing with the example of a firm deciding which market to enter, the firm's manager may have a long experience with one of the two markets, while he may not be better equipped than board directors to assess the outcome of the entry into the other market.

The purpose of this Section is to characterize how the optimal contract depends on the precisions of the two signals received by the expert. We will refer to two cases. In the first, we will analyze situations in which the efficient decision is sensitive to both signals, α and β. Because the efficient decision is sensitive to α and β when the signals received by the expert have sufficiently similar precisions, we will refer to this case as one in which decisions have "similar complexity." In the second, we will analyze situations in which the efficient decision is sensitive to only one of the two signal, α or β. Because this happens when the signals received by the expert have sufficiently different precisions, we will refer to this case as one in which decisions have "different complexity."

5.1 Decisions of similar complexity

Consider the case in which the efficient decision is sensitive to α and β. By Lemmas 1, 2 and 3, under Assumptions 1 and 2 the efficient decision is A unless the expert observes $\left(\underline{\alpha}, \overline{\beta}\right)$, in which case the efficient decision is B.

Proposition 3 *Suppose that Assumptions 1 and 2 are satisfied and the efficient decision is sensitive to α and β, i.e., $\varepsilon_A \in [\underline{\varepsilon}_A(\varepsilon_B), \overline{\varepsilon}_A(\varepsilon_B)]$. Then, there exist $\underset{\sim A}{\varepsilon}(\varepsilon_B)$ and $\widetilde{\varepsilon}_A(\varepsilon_B)$ satisfying $\underline{\varepsilon}_A(\varepsilon_B) < \underset{\sim A}{\varepsilon}(\varepsilon_B) < \varepsilon_B < \widetilde{\varepsilon}_A(\varepsilon_B) < \frac{1}{2}$ such that:*

1. *If* $\underline{\varepsilon}_A(\varepsilon_B) \leq \varepsilon_A \leq \underset{\sim}{\varepsilon}_A(\varepsilon_B)$, *the optimal contract* w^* *satisfies:*

$$w^*_{AS} < w^*_{BS};$$

2. *If* $\underset{\sim}{\varepsilon}_A(\varepsilon_B) < \varepsilon_A \leq \widetilde{\varepsilon}_A(\varepsilon_B)$, *the optimal contract* w^* *satisfies:*

$$w^*_{AF} < w^*_{BF} < w^*_{AS} \leq w^*_{BS};$$

3. *If* $\widetilde{\varepsilon}_A(\varepsilon_B) < \varepsilon_A \leq \widehat{\varepsilon}_A(\varepsilon_B)$, *the optimal contract* w^* *satisfies:*

$$w^*_{AF} < w^*_{BF} < w^*_{BS} < w^*_{AS}.$$

Therefore, when signals have very similar precision (case 2) the optimal contract has the same form as the one described in Proposition 2 which describes the case in which the signals have identical precisions. However, as the precision of one of the signals increases relative to the other, the contract can cease to be monotonic in the decisions, that is, which decision commands a higher pay depends on the realization of profits.

5.2 Decisions with different levels of complexity/uncertainty

Consider the case in which the efficient decision is sensitive to only one of the two signals, α or β.

Proposition 4 *Suppose that Assumptions 1 and 2 are satisfied. (1) If the efficient decision is sensitive to* α, *the optimal contract satisfies*

$$w_{AF} < w_{BF} = w_{BS} < w_{AS}.$$

(2) If the efficient decision is sensitive to β, *the optimal contract satisfies*

$$w_{BF} < w_{AF} = w_{AS} < w_{BS}.$$

As in previous cases, the optimal contract is weakly increasing in performance to ensure that the expert makes the efficient decision contingent on the signal received. However, when the efficient decision is sensitive to only of the signals, pay is independent of profits if the expert makes the more complex decision. This result implies that we may observe pay that is unresponsive to performance, even when the contract is optimally set. Another feature of optimal contracts is that, contrary to the previous cases, the form of the contract depends on which decision is more complex and not on which decision is ex ante efficient. Even if in both cases contemplated in Proposition 4the ex ante efficient decision is A (Assumption 1) the ranking of salaries depends only on the precision with which the outcomes of the two decisions can be forecasted by the expert. It is important to note that the expert's pay is not monotonic in the decision made, that is, it is not the case that, contingent on returns, one of the decisions always leads to a higher pay.

To understand part 1 of Proposition 4, consider for simplicity that the signal α allows the expert to perfectly forecast the outcome of decision A ($\varepsilon_A = 0$). For the sake of the argument, let us first consider, as we did with previous cases, the ranking of public histories that would be optimal if our only concern were to elicit effort provision from the expert. If the expert always chose the efficient decision conditional on his information, then, an argument identical to the one made in Section 4 implies that the optimal contract (which induces effort at the least possible cost for the principal) would have the following ranking of salaries:

$$w_{AF} < w_{AS} < w_{BF} = w_{BS}$$

The obvious problem with such a contract is that the expert would choose B irrespectively of the signals he received. To avoid this, it is necessary to set $w_{AS} \geq w_{BF} = w_{BS}$, which leads to a ranking of salaries like the one specified in Part 1 of Proposition 4.

Consider now the symmetric case in which $\varepsilon_B = 0$ and $\varepsilon_A = 1/2$, i.e., the case the signal about decision A is uninformative and the signal about decision B is completely informative. As above, let us first analyze the form of the optimal contract if the expert always made the efficient decisions conditional on his information. Given that, in this case, an expert who shirks never plays B, the ex-ante inefficient decision, the likelihood of public history BS relative to any other history is higher for the expert when he exerts effort than when he does not (in fact it is infinitely higher). However, in contrast to the previous case, the same is not true for public history BF, because this public history has zero probability with either level of effort. In this situation, the optimal contract would specify that the expert's pay after public history BS should be higher than after any other public history. Since public history BF has zero probability independently of whether the expert exerts effort, any w_{BF} would be optimal. Set w_{BF} sufficiently low, since this would be optimal if we had to provide incentives to the expert to select the efficient decision. Now, our assumption that $\varepsilon_A = 1/2$ implies that the likelihood of public history AS relative to public history AF is the same for either level of effort (that is, contingent on making decision A, the probability of success is the same irrespectively of whether the expert exerted effort or not). Therefore, to minimize the risk faced by the expert, the optimal contract should specify $w_{AF} = w_{AS}$. Therefore, the optimal contract—if we did not have to provide incentives for the expert to make the efficient decision contingent on his information—would have:

$$w_{BF} < w_{AF} = w_{AS} < w_{BS}$$

Consider now the incentives that the above contract would give to an expert who is free to choose decision A or B. Unlike in the case in which the outcome of decision A is perfectly predictable, the expert now has the right incentives to make the efficient decision: when the expert exerts effort and predicts success for B, he prefers to play B ($w_{BS} > w_{AF} = w_{AS}$), and when he predicts failure for B, he prefers to play A ($w_{BF} < w_{AF} = w_{AS}$). Therefore, in this case, incentive compatibility constraints having to do with the expert's decisions do not lead to a re-ranking of salaries as compared to a situation in which those incentive compatibility constraints can be ignored.

6 Correlated outcomes

Until now we have considered the case in which A and B are independent. But in many cases the returns of the two decisions may not be independent. For example, if the expert is a manager who has to decide whether to launch a new product in one of two countries which are not too dissimilar, it is reasonable to expect that if the product is good, returns will tend to be large in both countries and if the product is bad profits will tend to be low in both. In these cases r_A and r_B may be positively correlated. In other cases, however, r_A and r_B may be negatively correlated. For example, if the expert is a manager who has to decide whether to adopt one of two technological standards, network externalities associated with standards imply that the returns from adopting the two standards are negatively correlated: if standard

A succeeds r_A will tend to be large and r_B will tend to be low, and vice versa. As is obvious, when the returns from the two decisions have a positive correlation, obtaining information about their likely realizations is unlikely to have an impact on the efficient decision and the value of an expert that can obtain this information is lower. For this reason, in the following we focus on the case in which the two probability distributions have negative correlation and for the sake if simplicity we consider the case in which there exists negative perfect correlation between A and B. We assume in other words that with probability a, $(r_A, r_B) = (S, F)$, and with probability $1 - a$, $(r_A, r_B) = (F, S)$. Notice that the equivalent of Assumption 1 that guarantees that A is ex-ante efficient is that $a > \frac{1}{2}$.

To make the results comparable with the previous section we assume that if the expert exerts effort, he obtains only one signal, α rather than two. The reason is that this implies that, as in the previous section, the expert receives a signal with only two possible intensities *for each of the two projects*. For example, for A $\overline{\alpha}$ is favorable and $\underline{\alpha}$ is unfavorable; the opposite holds for B. With two independent signals, the perfect negative correlation of the returns of A and B would imply that the expert receives a signal for with four different intensities *for each of the two projects*. For example, for A, $\left(\overline{\alpha}, \underline{\beta}\right)$ is very favorable, $\left(\underline{\alpha}, \overline{\beta}\right)$ is very unfavorable and $\left(\overline{\alpha}, \overline{\beta}\right)$ and $\left(\underline{\alpha}, \underline{\beta}\right)$ are intermediate; the opposite holds for B.

The conditional probability distribution of signal α is

$$\Pr\left(\alpha = \overline{\alpha} \mid (r_A, r_B)\right) = \begin{cases} 1 - \varepsilon & \text{if } (r_A, r_B) = (S, F) \\ \varepsilon & \text{if } (r_A, r_B) = (F, S) \end{cases},$$

with $\varepsilon \in \left[0, \frac{1}{2}\right)$. As in the previous sections, if the expert shirks, he receives no signal.

In the following we will say that the efficient decision is sensitive to α if the efficient decision is not a constant function of the realization of α and we will say that it is insensitive to α otherwise. As is obvious we only consider the case in which the efficient decision is sensitive to α (or otherwise the signal would have no value). As is also obvious, if the efficient decision is sensitive to α, the efficient decision is A conditional on $\overline{\alpha}$ and B conditional on $\underline{\alpha}$.

The following proposition provides a characterization of the optimal contract for this information structure:

Proposition 5 *Consider the case of perfect negative correlation of A and B. Suppose that A is ex-ante efficient and that the efficient decision is sensitive to α. Then the optimal contract* *satisfies*

$$w_{AF} < w_{BF} \leq w_{AS} < w_{BS}$$

Therefore, the optimal contract yields the same ranking of salaries as in the case in which decisions are independent and signal precisions are sufficiently similar (compare with Proposition 2 and with part 2 of Proposition 3). The rationale for the contract form is the same as in that case. On the one hand, monotonicity with respect to returns guarantees that if the expert exerts effort, he will make the decision that is efficient contingent on the signal received. On the other hand, paying the expert more if he makes the ex ante inefficient decision encourages him to exert effort.

7 Uninformative prior

The results in Propositions 2-4 are derived under the assumption that one of the decisions is ex ante efficient (Assumption 1). In this section we want to briefly discuss the extent to which our results depend on this assumption. To do this we make the extreme assumption that $a = b$. In words, we make the assumption that the priors are so uninformative that the two decisions are equally attractive ex-ante. For the sake of simplicity we omit the proof of the results and we simply relate them to the results of the previous sections.

Consider first the case in which decisions are independent.

Proposition 1 contemplates the case of extreme symmetry, i.e., $a = b$, and $\varepsilon_A = \varepsilon_B$, and states that in this case the optimal contract conditions pay only on performance.

Consider now the case in which $\varepsilon_A \neq \varepsilon_B$. Notice that $a = b$ implies that $\underline{\varepsilon}_A(\varepsilon_B) = \overline{\varepsilon}_A(\varepsilon_B) = \varepsilon_B$. This implies that the efficient decision is sensitive to α if and only if $\varepsilon_A < \varepsilon_B$ and the efficient decision is sensitive to β if and only if $\varepsilon_A > \varepsilon_B$. In other words when $\varepsilon_A \neq \varepsilon_B$, the case contemplated in Proposition 3, i.e., the case that the efficient decision depends on both signals, is impossible. The optimal contract for this case leads to the same ranking of salaries as the one discussed in Proposition 4

$$w_{AF} < w_{BS} = w_{BF} < w_{AS}.$$

This is not surprising, since Proposition 4 shows that the efficient decision is sensitive to only one signal, the ranking of salaries does not depend on which decision is ex ante efficient, but only on the precisions of the signals received by the expert.

Consider now the case in which return realizations of the two decisions are perfectly negatively correlated and in which $a = b$. In this case, the ex ante symmetry carries through to the optimal contract that would make the expert's pay depend only on the return realization (S or F) and not also on his decision (A or B).

The previous discussion serves to reinforce the idea that uncertainty about the efficient decisions should translate into contracts that condition pay only on the principal's objective (returns, in our case) exclusively in cases of extreme symmetry in which the priors on the two decisions are identical ($a = b$), and the precision of the signals received by the expert are also identical.

8 Discussion and Applications

8.1 Implications for Contract Design

The model presented here has several implications for the design of executive compensation contracts:

1. *Pay should generally depend on observable decision and not only on returns.* When return-relevant decisions are observable and the efficient return-relevant decision is not known by shareholders, then it is generally optimal to make the manager's pay depend on the observable decision. Pay should not depend on the observable decision only in the extreme case in which shareholders (or board directors) are unable to rank decisions ex ante and the set of available decisions or the manager's idiosyncratic features are such that the manager would receive information with the same degree of precision about all feasible decisions.

2. *Expectations about the efficient decision matter.* The relation between pay and observable decisions may depend on which decision shareholders *expect* to be efficient, even if shareholders (or board directors) *do not know* with certainty which decision is efficient.

3. *Optimal contracts may pay managers to go against the (ex ante) odds.* This implication has the same statistical inference logic of the informativeness principle. If managers are not specialized, then they will go against the ex ante odds only if they have exerted effort and obtained information that favors the ex ante inefficient decision. Therefore, to elicit

effort in information acquisition, it may be optimal to explicitly reward decisions that go against shareholders' expectations.

4. *The precision of information matters.* A manager may be able to predict the consequences of some decisions better than those of others. This may be due to his own idiosyncratic specialization or may occur because the outcomes of some decisions are inherently more difficult to forecast. In these cases the ranking of salaries is determined by the precision of information and not by shareholders' ex-ante expectations on the efficient decision.

5. *Lack of performance-sensitivity may be optimal following some decisions.* When the information a manager obtains about a decision is sufficiently less precise than the information he obtains on the alternative, his pay should be respond to performance if he makes the decision on which he obtains more precise information. By contrast if a manager makes a decision on which he has comparatively sufficiently less precise information he should not be rewarded for success or penalized for failure.

6. *Pay may not be monotonic in decisions.* This occurs in all cases in which the manager obtains information with sufficiently different degrees of precision on the feasible decisions. In these cases, whether the manager is paid more for making one decision or the other will depend on performance: conditional on low performance, pay will be higher when the manager makes the decision on which he has less precise information; conditional on high performance, however, pay will be higher when the manager takes the decision in which he has more precise information.

There are, of course, several factors, which we have not explicitly considered in this paper, that should be taken into account when applying the above insights to the design of actual compensation contracts. First, making compensation depend on managerial decisions may induce distortions when some decisions are not observable. Second, we have assumed that the ex ante distribution of returns was common knowledge. If this is not the case, there may be disagreement among principals about the form of the compensation contract (while there would be no disagreement about rewarding the manager for success and penalizing him for failure). Further, shareholders' prior may differ from the manager's, which may make incentives backfire. Finally, making pay depend on decisions opens the door to ad hoc justifications for managerial pay excesses if the board serves the interests of the CEO.

8.2 Pay-Performance Sensitivity

Although conclusions depend on the metric used to measure it (se e.g., Murphy, 1999; Hall and Liebman, 1998; Baker and Hall, 2004), CEOs' pay-performance sensitivity is often believed to be too low. Our results may provide an explanation for this perceived lack of performance sensitivity. As discussed above, when managerial specialization or inherent forecasting complexity make it harder to forecast the realization of one decision relative to its alternative, it is optimal to make pay unresponsive to performance when the expert takes the former decision. To the extent that pay is explicitly or implicitly tied to CEOs' decisions as predicted by our model, the asymmetry of information on different decisions could lead to a relatively high incidence of low performance sensitivity. For example, if $a = 0.6$, $b = 0.5$, $\varepsilon_A = 0.3$, and $\varepsilon_B = 0.4$ zero performance sensitivity would occur with probability of 0.46.

8.3 Mergers and Acquisitions

Managers often receive large bonuses when they acquire other firms (Grinstein and Hribar, 2004), even if acquisitions have been shown to reduce, on average, the acquiring firms' value (Andrade et al., 2001; Moeller et al., 2004 and 2005). If acquisitions reduce acquiring firms' value, it seems puzzling that the managers of acquiring firms are rewarded for acquisitions. Several possible explanations for this puzzle have been proposed. The first one is, simply, that the acquisition process takes extraordinary effort from managers, so they need to be paid more in the event of an acquisition to induce them to put that extra effort or, simply, to satisfy their ex post participation constraint. Although the length of the acquisition process seems to increase the manager's bonus, the explanation based on effort is, at best, partial (Grinstein and Hribar, 2004). Further, other authors have proposed that managers have intrinsic incentives to acquire other firms because they are motivated by empire building (Jensen, 1986) or that they tend to overestimate their ability to benefit from the acquisition (the hubris hypothesis of Roll, 1986). According to these authors, incentives would, in fact, be needed, to dissuade managers from acquiring other firms. Another solution to the puzzle is that acquisitions, despite lowering the stock price of the acquiring firm, are optimal for the acquiring firm's shareholders (Jovanovic and Braguinsky, 2004). This may be the case if the acquisition, despite being efficient, reveals negative information to the market about the acquiring firm. Proponents of this view, however, have not discussed whether managerial pay should be explicitly tied to acquisitions. Finally, acquisition bonuses may be the result of managers' discretion over their own compensation. Thus, managers may use acquisitions as a way to justify pay increases to investors (Bebchuk and Fried, 2004, ch. 10).

Our model provides an alternative explanation for the relation between managers' pay and performance. First, it shows that, irrespectively of the sign of the effect of acquisitions on firm value, acquisitions should generally be taken into account when determining CEOs' pay. Second, it provides an optimal contracting explanation for the apparently perverse relation between acquisitions' negative impact on acquiring firms' value and managerial bonuses: it is precisely because acquisitions *on average* reduce the acquiring firms' value that managers should be rewarded for acquiring other firms. Otherwise, managers may refrain from undertaking value-increasing acquisitions. In terms of the testable predictions that follow from the model, it is worth noting that our results imply that the manager of an acquiring firm whose acquisition fails should be penalized for it, and should earn less than the manager of a successful—comparable—firm who had not carried out and acquisition.

9 Conclusion

In this paper, we have studied an agency problem in which a principal hires an expert agent to make a decision that determines the principal's profits. The reason why the principal hires the expert is that she does not know which is the efficient decision, while the expert's superior skills allow him, if he exerts effort, to obtain more precise information about the efficient decision. We consider a context in which the principal observes both the expert's decision and the resulting profits, but not the effort exerted by the expert to obtain information. Therefore, the principal can condition the expert's pay both on profits and on the observable decision. In this context, the expert's contract has to solve two related incentive problems: it has to a) provide incentives to the expert to exert effort to acquire information and b) ensure that the expert has the incentives to make the efficient decision conditional on his information.

The assumptions of the model are meant to capture what we consider to be the main agency problem between shareholders and managers. Shareholders and, certainly, board directors—who actually set the manager's compensation—can observe the most relevant decisions made by managers, namely those firm-wide strategic decisions that determine the firm's scope, organizational form, technology, financing and the markets in which it operates. Therefore, they can, in principle, condition the manager's pay on those strategic decisions. The problem is, however, that the very reason why they delegate those decisions to an expert manager is, most often, that they believe that the manager is better equipped than they are to make efficient decisions for the firm.

We study the optimal contract in this context and obtain several novel predictions about the form of the executive compensation contract. First, we show that, under very general conditions, the optimal contract conditions pay not only on profits, but also on the decision made by the manager. Although similar in spirit, this result does not follow from Holmström's (1979) informativeness principle. Second, we show that shareholders' (or board directors') ex ante beliefs about the efficient decision determine the form of the optimal compensation contract, a prediction that, to our knowledge, had not been obtained before in agency models. Third, we show that degrees of precision with which an expert manager learns about different decisions also influences the form of the compensation contract. Fourth, we characterize the relation between profits, decisions, and pay implied by the optimal contract in different contexts and find, among other implications, that: i) it may optimal to reward managers for making decisions which are ex ante inefficient; ii) it may be optimal to shield managers' compensation from performance measures; and iii) optimal contracts may be nonmonotonic in decisions in the sense that the decision that is rewarded conditional on a success may be different from the decision that is rewarded conditional on failure.

We have discussed how our predictions may shed light on compensation practices, such as rewarding CEOs for acquiring other firms or for reductions in employment. We believe, however, that the normative implications of our model have a much broader application, since they provide guidelines, beyond the informativeness principle, concerning the variables that should be included, and in which way, in executive compensation contracts. We also believe that our results can also be applied to the study of the compensation contracts of expert agents other than firm executives, such as medical doctors, investment analysts or lawyers.

References

[1] Andrade G., M. Mitchell, and E. Stafford (2001) "New Evidence and Perspectives on Mergers," *The Journal of Economic Perspectives*, 15, 103–120.

[2] Ang, J. and P.Y. Keys "An Analysis of Boards' Actions Towards CEOs During Periods of Employee Layoffs", mimeo.

[3] Baker, G. P. and J. Hall (2004) "CEO Incentives and Firm Size", *Journal of Labor Economics*, 22, 767-798.

[4] Bebchuk, L. and J. Fried (2004). *Pay without Performance: The Unfulfilled Promise of Executive Compensation.* Harvard University Press.

[5] Billger, S.M. and K.F. Hallock (2005) "Mass Layoffs and Management Turnover", *Industrial Relations* 44, 463-489.

[6] Caruana, G. and M. Celentani (2002) "Career Concerns and Contingent Compensation" CEMFI Working Paper 205.

[7] Grinstein Y., and Hribar P. (2004). "CEO Compensation and Incentives: Evidence from M & A Bonuses." *Journal of Financial Economics*, 71, 119–143.

[8] Hall, B.J. and Liebman, J.B. (1998), "Are CEOs Really Paid Like Bureaucrats?", *The Quarterly Journal of Economics*, 113, ,653–691.

[9] Hallock, K.F. (1998) "Layoffs, Top Executive Pay, and Firm Performance" *American Economic Review*, 88-4, 711-723.

[10] Holmström, B. (1979) "Moral Hazard and Observability", *Bell Journal of Economics*, 10, 231-259.

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

[12] Jensen, M.C. (1986) "Agency Costs of Free Cash Flow, Corporate Finance, and Takeovers," *American Economic Review* 76, 323—329.

[13] Jensen, M. and W. Meckling (1976) "Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure," *Journal of Financial Economics*, 3, 305–360.

[14] Jovanovic, B. and S. Braguinsky (2004) "Bidder Discounts and Target Premia in Takeovers," *American Economic Review*, 94, 46–56.

[15] Moeller S., F. Schlingemann, and R. Stultz (2005) "Wealth Destruction on a Massive Scale? A Study of Acquiring-Firm Returns in the Recent Merger Wave," *Journal of Finance*, 60, 757–782.

[16] Moeller S., F. Schlingemann, and R. Stultz (2004) "Firm Size and the Gains from Acquisitions," *Journal of Financial Economics*, 73, 201-228.

[17] Murphy, K.J. (1999) "Executive Compensation" in O. Ashenfelter and D. Card (1999) (eds.) *Handbook of Labor Economics*, Vol. 3, North Holland, Amsterdam.

[18] Roll, R. (1986) "The Hubris Hypothesis of Corporate Takeovers," *Journal of Business* 59, 197—216.

A Proofs

A.1 Proof of Lemma 1

Immediate from observing that 1 implies that $\overline{a} > \underline{b}$.

A.2 Proof of Lemma 2

Notice that $\hat{d}\left(\underline{\alpha}, \overline{\beta}\right) = B$ is equivalent to $\underline{a} < \overline{b}$. From the definitions of $\underline{a}$ and $\overline{b}$ we obtain

$$a\varepsilon_A(b(1-\varepsilon_B) + (1-b)\varepsilon_B) \leq b(1-\varepsilon_B)(a\varepsilon_A + (1-a)(1-\varepsilon_A))$$

which is equivalent to

$$\varepsilon_A < \frac{(1-a)\, b\,(1-\varepsilon_B)}{a\,(1-b)\,\varepsilon_B + (1-a)\, b\,(1-\varepsilon_B)} = 1 - \overline{\varepsilon}_A\left(\varepsilon_B\right).$$

A.3 Notation

We denote by $\pi_{\alpha\beta}$ the unconditional probability of the expert receving signals (α, β) if he exerts effort. By the assumption of independence of A and B and of α and β, we have

$$\pi_{\alpha\beta} = \Pr\left(\alpha, \beta\right) = \Pr\left(\alpha\right) \Pr\left(\beta\right) \tag{6}$$

We denote by $V_{\overline{A}}$ the expected utility of the salary payment to the expert when he chooses A after having observed a favorable signal on A, i.e. after having observed signal $\overline{\alpha}$. In a similar way we denote by $V_{\underline{A}}$ the expected utility of the salary payment to the expert when he chooses A after having observed an unfavorable signal on A, i.e. after having observed signal $\underline{\alpha}$. Notice that the assumption of independence of A and B and of α and β implies that the expected utility of the salary payment to the expert when he chooses A depends on α, but is independent of β. We define $V_{\overline{B}}$ and $V_{\underline{B}}$ in a similar way.

We denote by V_A (respectively, V_B) the unconditional expected utility of the salary payment to the expert when he chooses A (respectively, B). We therefore have

$$V_A = \pi_{\overline{\alpha}\underline{\beta}} V_{\overline{A}} + \pi_{\underline{\alpha}\overline{\beta}} V_{\underline{A}} + \pi_{\overline{\alpha}\overline{\beta}} V_{\overline{A}} + \pi_{\underline{\alpha}\underline{\beta}} V_{\underline{A}}$$
$$V_B = \pi_{\overline{\alpha}\underline{\beta}} V_{\underline{B}} + \pi_{\underline{\alpha}\overline{\beta}} V_{\overline{B}} + \pi_{\overline{\alpha}\overline{\beta}} V_{\overline{B}} + \pi_{\underline{\alpha}\underline{\beta}} V_{\underline{B}}$$

Finally we denote by V the expected utility of the expert of a contract, under the assumption that the expert exerts effort and makes the efficient decision. In the case in which the efficient decision is sensitive to α and β (which implies that the efficient decision is B if and only if $(\alpha, \beta) = (\underline{\alpha}, \overline{\beta})$), for example we have

$$V \equiv \pi_{\overline{\alpha}\underline{\beta}} V_{\overline{A}} + \pi_{\underline{\alpha}\overline{\beta}} V_{\overline{B}} + \pi_{\overline{\alpha}\overline{\beta}} V_{\overline{A}} + \pi_{\underline{\alpha}\underline{\beta}} V_{\underline{A}} - g \tag{7}$$

A.4 Proof of Propositions 1, 2, and 3

We start by characterizing the solution to the cost minimization problem when the efficient decision is sensitive to α and β. Recall that when the efficient decision is sensitive to α and β, the efficient decision is B if and only if $(\alpha, \beta) = (\underline{\alpha}, \overline{\beta})$. The contract that minimizes the salary cost of inducing effort and implementing the efficient decision is therefore a solution to the following problem

$$\begin{aligned}
\min_{w \in \mathbb{R}^4} \quad & \pi_{\overline{\alpha}\underline{\beta}} \left(\overline{a} w_{AS} + (1-\overline{a}) w_{AF}\right) + \pi_{\underline{\alpha}\overline{\beta}} \left(\overline{b} w_{BS} + (1-\overline{b}) w_{BF}\right) + \\
& + \pi_{\overline{\alpha}\overline{\beta}} \left(\overline{a} w_{AS} + (1-\overline{a}) w_{AF}\right) + \pi_{\underline{\alpha}\underline{\beta}} \left(\underline{a} w_{AS} + (1-\underline{a}) w_{AF}\right) \\
\text{s.t.} \quad & \pi_{\overline{\alpha}\underline{\beta}} V_{\overline{A}} + \pi_{\underline{\alpha}\overline{\beta}} V_{\overline{B}} + \pi_{\overline{\alpha}\overline{\beta}} V_{\overline{A}} + \pi_{\underline{\alpha}\underline{\beta}} V_{\underline{A}} - g \geq \overline{U} && (8) \\
& \pi_{\overline{\alpha}\underline{\beta}} V_{\overline{A}} + \pi_{\underline{\alpha}\overline{\beta}} V_{\overline{B}} + \pi_{\overline{\alpha}\overline{\beta}} V_{\overline{A}} + \pi_{\underline{\alpha}\underline{\beta}} V_{\underline{A}} - g \geq \pi_{\overline{\alpha}\underline{\beta}} V_{\overline{A}} + \pi_{\underline{\alpha}\overline{\beta}} V_{\underline{A}} + \pi_{\overline{\alpha}\overline{\beta}} V_{\overline{A}} + \pi_{\underline{\alpha}\underline{\beta}} V_{\underline{A}} && (9) \\
& \pi_{\overline{\alpha}\underline{\beta}} V_{\overline{A}} + \pi_{\underline{\alpha}\overline{\beta}} V_{\overline{B}} + \pi_{\overline{\alpha}\overline{\beta}} V_{\overline{A}} + \pi_{\underline{\alpha}\underline{\beta}} V_{\underline{A}} - g \geq \pi_{\overline{\alpha}\underline{\beta}} V_{\underline{B}} + \pi_{\underline{\alpha}\overline{\beta}} V_{\overline{B}} + \pi_{\overline{\alpha}\overline{\beta}} V_{\overline{B}} + \pi_{\underline{\alpha}\underline{\beta}} V_{\underline{B}} && (10) \\
& V_{\overline{A}} \geq V_{\underline{B}} && (11) \\
& V_{\overline{B}} \geq V_{\underline{A}} && (12) \\
& V_{\overline{A}} \geq V_{\overline{B}} && (13) \\
& V_{\underline{A}} \geq V_{\underline{B}} && (14)
\end{aligned}$$

LEMMA 4 (12) *is not binding.*

Proof. We can rewrite (9):

$$\pi_{\underline{\alpha}\overline{\beta}}(V_{\overline{B}} - V_{\underline{A}}) \geq g,$$

which implies that $V_{\overline{B}} > V_{\underline{A}}$. ■

LEMMA 5 $w_{AS} > w_{AF}$.

Proof. Lemma 4 and (13) imply $V_{\overline{A}} \geq V_{\overline{B}} > V_{\underline{A}}$. Thus,

$$V_{\overline{A}} > V_{\underline{A}}. \tag{15}$$

Since $\overline{a} > \underline{a}$, this requires $w_{AS} > w_{AF}$. ■

LEMMA 6 (11) *is not binding.*

Proof. (15) and (14) imply $V_{\overline{A}} > V_{\underline{B}}$. ■

Given Lemmas 4 and 6, the first order conditions are:

$$\frac{1}{U'(w_{AS})} = \lambda_P + \lambda_B + \frac{\left(\lambda_4 - \lambda_A \pi_{\underline{\alpha}\overline{\beta}}\right)\underline{a} + \lambda_3 \overline{a}}{\left(\pi_{\overline{\alpha}\underline{\beta}} + \pi_{\overline{\alpha}\overline{\beta}}\right)\overline{a} + \pi_{\underline{\alpha}\underline{\beta}}\underline{a}}; \tag{16}$$

$$\frac{1}{U'(w_{AF})} = \lambda_P + \lambda_B + \frac{\left(\lambda_4 - \lambda_A \pi_{\underline{\alpha}\overline{\beta}}\right)(1-\underline{a}) + \lambda_3 (1-\overline{a})}{\left(\pi_{\overline{\alpha}\underline{\beta}} + \pi_{\overline{\alpha}\overline{\beta}}\right)(1-\overline{a}) + \pi_{\underline{\alpha}\underline{\beta}}(1-\underline{a})}; \tag{17}$$

$$\frac{1}{U'(w_{BS})} = \lambda_P + \lambda_A - \frac{\lambda_B\left[\left(\pi_{\overline{\alpha}\underline{\beta}} + \pi_{\underline{\alpha}\underline{\beta}}\right)\underline{b} + \pi_{\overline{\alpha}\overline{\beta}}\overline{b}\right] + \lambda_3 \overline{b} + \lambda_4 \underline{b}}{\pi_{\underline{\alpha}\overline{\beta}}\overline{b}}; \tag{18}$$

$$\frac{1}{U'(w_{BF})} = \lambda_P + \lambda_A - \frac{\lambda_B\left[\left(\pi_{\overline{\alpha}\underline{\beta}} + \pi_{\underline{\alpha}\underline{\beta}}\right)(1-\underline{b}) + \pi_{\overline{\alpha}\overline{\beta}}\left(1-\overline{b}\right)\right] + \lambda_3\left(1-\overline{b}\right) + \lambda_4(1-\underline{b})}{\pi_{\underline{\alpha}\overline{\beta}}\left(1-\overline{b}\right)}. \tag{19}$$

LEMMA 7 $w_{BS} > w_{BF}$.

Proof. Lemma 4 and (14) imply $V_{\overline{B}} > V_{\underline{A}} \geq V_{\underline{B}}$. Thus,

$$V_{\overline{B}} > V_{\underline{B}}.$$

Since $\overline{b} > \underline{b}$, this requires $w_{BS} > w_{BF}$. ■

LEMMA 8 $w_{AS} > w_{BF}$ *and* $w_{BS} > w_{AF}$.

Proof. Suppose that $w_{AS} \leq w_{BF}$. This implies that $w_{AF} < w_{AS} \leq w_{BF} < w_{BS}$ and contradicts (11), (13), and (14). Suppose that $w_{BS} \leq w_{AF}$. This implies that $w_{BF} < w_{BS} \leq w_{AF} < w_{AS}$ and contradicts (12). ■

LEMMA 9 (9) *is binding.*

Proof. Suppose (9) is not binding at an optimum. This implies that

$$\pi_{\underline{\alpha}\overline{\beta}}(V_{\overline{B}} - V_{\underline{A}}) > g.$$

Consider the following local deviation: $dw = (0, \Delta_{AF}, \Delta_{BS}, 0)$, such that $\Delta_{AF} > 0$, $\Delta_{BS} < 0$ and $\mathbf{D}V dw = 0$ (i.e., the change keeps V unchanged):

$$\pi_{\overline{\alpha}\underline{\beta}}U'(w_{AF})(1-\overline{a})\Delta_{AF}+\pi_{\underline{\alpha}\overline{\beta}}U'(w_{BS})\overline{b}\Delta_{BS}+\pi_{\overline{\alpha}\overline{\beta}}U'(w_{AF})(1-\overline{a})\Delta_{AF}+\pi_{\underline{\alpha}\underline{\beta}}U'(w_{AF})(1-\underline{a})\Delta_{AF} = 0$$

i.e.,

$$\Delta_{BS} = -\Delta_{AF}\left(\frac{U'(w_{AF})}{\pi_{\underline{\alpha}\overline{\beta}}\overline{b}U'(w_{BS})}\right)[(\pi_{\overline{\alpha}\underline{\beta}} + \pi_{\overline{\alpha}\overline{\beta}})(1-\overline{a}) + \pi_{\underline{\alpha}\underline{\beta}}(1-\underline{a})].$$

By construction, (8) and (10) hold, since the change does not affect V and reduces V_B. If the change is small enough (9), (11) and (12) hold (recall that (11) and (12) have been shown to be nonbinding and (9) is nonbinding by hypothesis). (13) and (14) hold since the change increases $V_{\overline{A}}$ and $V_{\underline{A}}$ and decreases $V_{\overline{B}}$ and $V_{\underline{B}}$. The change in expected cost is:

$$\begin{aligned}&\pi_{\overline{\alpha}\underline{\beta}}(1-\overline{a})\Delta_{AF} + \pi_{\underline{\alpha}\overline{\beta}}\overline{b}\Delta_{BS} + \pi_{\overline{\alpha}\overline{\beta}}(1-\overline{a})\Delta_{AF} + \pi_{\underline{\alpha}\underline{\beta}}(1-\underline{a})\Delta_{AF} = \\ &= \Delta_{AF}[(\pi_{\overline{\alpha}\underline{\beta}} + \pi_{\overline{\alpha}\overline{\beta}})(1-\overline{a}) + \pi_{\underline{\alpha}\underline{\beta}}(1-\underline{a})]\left(1 - \frac{U'(w_{AF})}{U'(w_{BS})}\right) < 0, \qquad (20)\end{aligned}$$

with the inequality following from $w_{BS} > w_{AF}$ and the concavity of U. This contradicts the hypothesis that (9) is nonbinding. ■

LEMMA 10 *If* (10) *is not binding,* (14) *is binding.*

Proof. Assume that (14) is not binding. From (18) and (19) we obtain

$$\frac{1}{U'(w_{BS})} = \lambda_P + \lambda_A - \frac{\lambda_3}{\pi_{\underline{\alpha}\overline{\beta}}};$$
$$\frac{1}{U'(w_{BF})} = \lambda_P + \lambda_A - \frac{\lambda_3}{\pi_{\underline{\alpha}\overline{\beta}}};$$

which implies $w_{BS} = w_{BF}$, and contradicts Lemma 7. ■

We now consider three cases:

- Case 1: $\pi_{\overline{\alpha}\underline{\beta}} > \pi_{\underline{\alpha}\overline{\beta}}$;
- Case 2: $\pi_{\overline{\alpha}\underline{\beta}} = \pi_{\underline{\alpha}\overline{\beta}}$;
- Case 3: $\pi_{\overline{\alpha}\underline{\beta}} < \pi_{\underline{\alpha}\overline{\beta}}$.

Lemmas 11 to 15 refer to Case 1:

LEMMA 11 (10) *is not binding at an optimum.*

Proof. Suppose it is binding. Since by Lemma 9 (9) is binding, it follows that $V_A = V_B$. Thus:

$$\pi_{\overline{\alpha}\underline{\beta}}(V_{\overline{A}} - V_{\underline{B}}) + \pi_{\overline{\alpha}\overline{\beta}}(V_{\overline{A}} - V_{\overline{B}}) + \pi_{\underline{\alpha}\underline{\beta}}(V_{\underline{A}} - V_{\underline{B}}) = \pi_{\underline{\alpha}\overline{\beta}}(V_{\overline{B}} - V_{\underline{A}}) \tag{21}$$

From (13) and (14)

$$\pi_{\overline{\alpha}\overline{\beta}}(V_{\overline{A}} - V_{\overline{B}}) + \pi_{\underline{\alpha}\underline{\beta}}(V_{\underline{A}} - V_{\underline{B}}) \geq 0. \tag{22}$$

From (21) and (22)

$$\pi_{\underline{\alpha}\overline{\beta}}(V_{\overline{B}} - V_{\underline{A}}) \geq \pi_{\overline{\alpha}\underline{\beta}}(V_{\overline{A}} - V_{\underline{B}}). \tag{23}$$

Given that $\pi_{\overline{\alpha}\underline{\beta}} > \pi_{\underline{\alpha}\overline{\beta}}$, from (23), $V_{\overline{B}} - V_{\underline{A}} > V_{\overline{A}} - V_{\underline{B}}$. But then, (13) implies $V_{\overline{B}} - V_{\underline{A}} > V_{\overline{B}} - V_{\underline{B}}$, which implies $V_{\underline{A}} < V_{\underline{B}}$ and a contradiction to (14) is obtained. ■

LEMMA 12 (14) *is binding at an optimum.*

Proof. Immediate from Lemmas 10 and 11. ■

Lemma 13 *Suppose* (13) *is binding at an optimum. The optimal contract satisfies:*

$$\begin{aligned} &For \quad \underline{\varepsilon}_A(\varepsilon_B) \leq \varepsilon_A \leq \underset{\sim}{\varepsilon}_A(\varepsilon_B), w_{BF} \leq w_{AF} < w_{AS} < w_{BS} \\ &For \quad \underset{\sim}{\varepsilon}_A(\varepsilon_B) < \varepsilon_A \leq \widetilde{\varepsilon}_A(\varepsilon_B), w_{AF} < w_{BF} < w_{AS} \leq w_{BS}; \\ &For \quad \widetilde{\varepsilon}_A < \varepsilon_A \leq \bar{\varepsilon}_A(\varepsilon_B), w_{AF} < w_{BF} < w_{BS} < w_{AS}. \end{aligned}$$

Proof. Define

$$\begin{aligned} \underset{\sim}{L} &= \overline{a}\underline{b} - \underline{a}\overline{b} \\ \widetilde{L} &= \left(\overline{a}\underline{b} - \underline{a}\overline{b}\right) + \left(\overline{b} - \underline{b}\right) - \left(\overline{a} - \underline{a}\right). \end{aligned}$$

Notice that $\underset{\sim}{L}$ and $\widetilde{L}$ are functions of ε_A and ε_B because the conditional probabilities of success are functions of ε_A and ε_B.

Recall from Lemma 12 that (14) is binding. From (13) and (14) with equality

$$U(w_{BS}) = \frac{\left[\underline{a}\left(1-\overline{b}\right)-\overline{a}\left(1-\underline{b}\right)\right]U(w_{BF})}{\overline{a}\underline{b}-\underline{a}\overline{b}} - \frac{\left[\underline{a}(1-\overline{a})-\overline{a}(1-\underline{a})\right]U(w_{AF})}{\overline{a}\underline{b}-\underline{a}\overline{b}}; \tag{24}$$

$$U(w_{AS}) = \frac{\left[\underline{b}\left(1-\overline{b}\right)-\overline{b}\left(1-\underline{b}\right)\right]U(w_{BF})}{\overline{a}\underline{b}-\underline{a}\overline{b}} - \frac{\left[(1-\overline{a})\underline{b}-(1-\underline{a})\overline{b}\right]U(w_{AF})}{\overline{a}\underline{b}-\underline{a}\overline{b}}. \tag{25}$$

Substituting (24) and (25) into (12) we obtain

$$\frac{U(w_{BF}) - U(w_{AF})}{\underset{\sim}{L}} < 0. \tag{26}$$

Given that $\underset{\sim}{L}$ is decreasing in ε_A and $\underset{\sim}{L} = 0$ when $\varepsilon_A = \underset{\sim}{\varepsilon}_A(\varepsilon_B)$, risk aversion implies that $\operatorname{sign}(w_{AF} - w_{BF}) = \operatorname{sign}\underset{\sim}{L}$.

From (24) and (25), $w_{BS} \geq w_{AS}$ if and only if

$$\widetilde{L} \times \frac{U(w_{BF}) - U(w_{AF})}{\underset{\sim}{L}} \geq 0. \tag{27}$$

By (26), (27) implies that $w_{BS} \geq w_{AS}$ if and only if $\widetilde{L} \leq 0$. Given that $\widetilde{L}$ is increasing in ε_A and $\widetilde{L} = 0$ when $\varepsilon_A = \widetilde{\varepsilon}_A(\varepsilon_B)$, risk aversion implies that $\operatorname{sign}(w_{AS} - w_{BS}) = \operatorname{sign}\widetilde{L}$. ∎

LEMMA 14 *If* (13) *is not binding, the optimal contract satisfies:*

$$\text{For} \quad \underline{\varepsilon}_A(\varepsilon_B) \leq \varepsilon_A \leq \underset{\sim}{\varepsilon}_A(\varepsilon_B) : w_{AS} < w_{BS};$$
$$\text{For} \quad \underset{\sim}{\varepsilon}_A(\varepsilon_B) < \varepsilon_A < \widetilde{\varepsilon}_A(\varepsilon_B) : w_{AF} < w_{BF} < w_{AS} < w_{BS}.$$

Proof. By Lemma 5, $w_{AS} > w_{AF}$. From (16), (17) and (18) this implies that $w_{BS} > w_{AS}$. From (14) with equality

$$U(w_{BS}) = \frac{\underline{a}U(w_{AS}) + (1-\underline{a})U(w_{AF}) - (1-\underline{b})U(w_{BF})}{\underline{b}}. \tag{28}$$

Substituting (28) into (13), (13) is not binding if and only if

$$U(w_{AS})\underset{\sim}{L} > \left(\underline{b} - \overline{b}\right)U(w_{BF}) + \left[\left(\overline{b} - \underline{b}\right) - \left(\underline{a}\overline{b} - \overline{a}\underline{b}\right)\right]U(w_{AF}). \tag{29}$$

If $\underset{\sim}{L} < 0$, condition (29) is equivalent to

$$-\frac{\overline{b} - \underline{b}}{\underset{\sim}{L}}U(w_{BF}) - \frac{\left(\underline{a}\overline{b} - \overline{a}\underline{b}\right) - \left(\overline{b} - \underline{b}\right)}{\underset{\sim}{L}}U(w_{AF}) > U(w_{AS}). \tag{30}$$

Substitute (28) into (12) and recalling from Lemma (4) that (12) is not binding we obtain

$$U(w_{AS}) > \frac{1}{\underline{a}}U(w_{BF}) - \frac{(1-\underline{a})}{\underline{a}}U(w_{AF}). \tag{31}$$

A necessary condition for (30) and (31) to both hold, is that the left hand side of (30) is greater than the right hand side of (31). Straightforward algebra shows that this inequality holds if and only if $w_{BF} > w_{AF}$. ■

LEMMA 15 *If* $\varepsilon_A \in [\widetilde{\varepsilon}_A(\varepsilon_B), \overline{\varepsilon}_A(\varepsilon_B)]$, (13) *is binding*

Proof. Notice first that for $\varepsilon_A \in [\widetilde{\varepsilon}_A(\varepsilon_B), \overline{\varepsilon}_A(\varepsilon_B)]$, $\underset{\sim}{L} < 0$. Suppose now that (13) is not binding. From Lemma 14, $w_{BS} > w_{AS}$. From (28),

$$U(w_{BS}) > U(w_{AS})$$

is equivalent to

$$U\left(w_{AS}\right) > \frac{(1-\underline{b})}{(\underline{a}-\underline{b})} U\left(w_{BF}\right) - \frac{(1-\underline{a})}{(\underline{a}-\underline{b})} U\left(w_{AF}\right). \tag{32}$$

Notice that since $\underset{\sim}{L} < 0$ (30) holds and a necessary condition for (32) to also hold is that the left hand side of (30) is greater than the right hand side of (32) and this is equivalent to

$$\widetilde{L}\left(U\left(w_{BF}\right) - U\left(w_{AF}\right)\right) < 0. \tag{33}$$

Notice now that in Lemma 14 it was established that $\underset{\sim}{L} < 0$ and (13) not binding imply $w_{BF} > w_{AF}$. This implies that a necessary condition for (33) to hold is $\widetilde{L} < 0$. But, because this requires $\varepsilon_A < \widetilde{\varepsilon}_A\left(\varepsilon_B\right)$, a contradiction is obtained. ■

Lemmas 16 and 17 refer to Case 2:

LEMMA 16 *Suppose* $\pi_{\overline{\alpha}\underline{\beta}} = \pi_{\underline{\alpha}\overline{\beta}}$. *If* (10) *is binding, then wages are as in Lemma 13.*

Proof. From (9) and (10) with equality and from $\pi_{\overline{\alpha}\underline{\beta}} = \pi_{\underline{\alpha}\overline{\beta}}$:

$$\left(\pi_{\overline{\alpha}\underline{\beta}} + \pi_{\overline{\alpha}\overline{\beta}}\right)\left(V_{\overline{A}} - V_{\overline{B}}\right) + \left(\pi_{\overline{\alpha}\underline{\beta}} + \pi_{\underline{\alpha}\underline{\beta}}\right)\left(V_{\underline{A}} - V_{\underline{B}}\right) = 0. \tag{34}$$

Given that (13) and (14) imply $\left(V_{\overline{A}} - V_{\overline{B}}\right) \geq 0$ and $\left(V_{\underline{A}} - V_{\underline{B}}\right) \geq 0$, (34) implies that (13) and (14) hold with equality. The rest of the proof follows along the lines of Lemma (13). ■

LEMMA 17 *Suppose* $\pi_{\overline{\alpha}\underline{\beta}} = \pi_{\underline{\alpha}\overline{\beta}}$. *If* (10) *is not binding, then wages are as in Lemma 14.*

Proof. If (10) is not binding, (14) is binding by Lemma 10. From $\pi_{\overline{\alpha}\underline{\beta}} = \pi_{\underline{\alpha}\overline{\beta}}$, (9) with equality and (10) with strict inequality we obtain

$$\left(\pi_{\overline{\alpha}\underline{\beta}} + \pi_{\overline{\alpha}\overline{\beta}}\right)\left(V_{\overline{A}} - V_{\overline{B}}\right) + \left(\pi_{\overline{\alpha}\underline{\beta}} + \pi_{\underline{\alpha}\underline{\beta}}\right)\left(V_{\underline{A}} - V_{\underline{B}}\right) > 0. \tag{35}$$

Because (14) is binding, $V_{\underline{A}} - V_{\underline{B}} = 0$ and from (35) we obtain $V_{\overline{A}} - V_{\overline{B}} > 0$, i.e., that (13) is not binding. The rest of the proof follows along the lines of Lemma 14. ■

Lemmas 18 to 21 refer to Case 3.

LEMMA 18 *Suppose* $\pi_{\overline{\alpha}\underline{\beta}} < \pi_{\underline{\alpha}\overline{\beta}}$. (13) *and* (14) *cannot be binding simultaneously.*

Proof. Suppose (13) and (14) are binding. Substituting them in (9) and (10), implies that (9) and (10) hold if and only if $\pi_{\overline{\alpha}\underline{\beta}} > \pi_{\underline{\alpha}\overline{\beta}}$ and a contradiction arises. ■

Lemma 19 *Suppose $\pi_{\overline{\alpha}\underline{\beta}} < \pi_{\underline{\alpha}\overline{\beta}}$. (1) If $a > 1/2$, $\varepsilon_A > \widetilde{\varepsilon}_A(\varepsilon_B)$. (2) If $a \leq 1/2$, $\varepsilon_A < \underset{\sim}{\varepsilon}_A(\varepsilon_B)$.*

Proof. If $a > \frac{1}{2}$, $\pi_{\overline{\alpha}\underline{\beta}} < \pi_{\underline{\alpha}\overline{\beta}}$ is equivalent to

$$\varepsilon_A > \frac{\Pr\left(\overline{\beta}\right) - a}{1 - 2a} > \widetilde{\varepsilon}_A(\varepsilon_B).$$

If $a \leq \frac{1}{2}$, $\pi_{\overline{\alpha}\underline{\beta}} < \pi_{\underline{\alpha}\overline{\beta}}$ is equivalent to

$$\varepsilon_A \leq \frac{\Pr\left(\overline{\beta}\right) - a}{1 - 2a} < \underset{\sim}{\varepsilon}_A(\varepsilon_B).$$

■

Lemma 20 *Suppose $\pi_{\overline{\alpha}\underline{\beta}} < \pi_{\underline{\alpha}\overline{\beta}}$. If (10) is not binding, then wages are as in Lemma 14.*

Proof. Same as in Lemma 17. ■

Lemma 21 *Suppose $\pi_{\overline{\alpha}\underline{\beta}} < \pi_{\underline{\alpha}\overline{\beta}}$. If (10) is binding the optimal contract satisfies*

$$\begin{aligned} For\ \underline{\varepsilon}_A(\varepsilon_B) &\leq \varepsilon_A < \underset{\sim}{\varepsilon}_A(\varepsilon_B) : w_{BF} < w_{AF} < w_{AS} < w_{BS}; \\ For\ \widetilde{\varepsilon}_A(\varepsilon_B) &< \varepsilon_A \leq \bar{\varepsilon}_A(\varepsilon_B) : w_{AF} < w_{BF} < w_{BS} < w_{AS}. \end{aligned}$$

Proof. From (9) and (10) binding we obtain

$$aU(w_{AS}) + (1-a)U(w_{AF}) = bU(w_{BS}) + (1-b)U(w_{BF}). \tag{36}$$

Assume first $a \leq \frac{1}{2}$. From (36) we obtain that

$$aU(w_{AS}) = bU(w_{BS}) + (1-b)U(w_{BF}) - (1-a)U(w_{AF}). \tag{37}$$

Substituting (37) into (12) and (14)

$$bU(w_{BS})\,G > \left[\varepsilon_A \Pr\left(\overline{\beta}\right) - \varepsilon_B \Pr\left(\underline{\alpha}\right)\right](1-b)\,U(w_{BF}) + \Pr\left(\overline{\beta}\right)(1-2\varepsilon_A)(1-a)\,U(w_{AF})\,;$$

$$bU(w_{BS})\,H \geq \left[(1-\varepsilon_B)\Pr\left(\underline{\alpha}\right) - \varepsilon_A \Pr\left(\underline{\beta}\right)\right](1-b)\,U(w_{BF}) - \Pr\left(\underline{\beta}\right)(1-2\varepsilon_A)(1-a)\,U(w_{AF})\,;$$

where

$$\begin{aligned} G &= (1-\varepsilon_B)\Pr\left(\underline{\alpha}\right) - \varepsilon_A \Pr\left(\overline{\beta}\right); \\ H &= \varepsilon_A \Pr\left(\underline{\beta}\right) - \varepsilon_B \Pr\left(\underline{\alpha}\right). \end{aligned}$$

Simple algebra shows that G is strictly positive and that $\pi_{\overline{\alpha}\underline{\beta}} < \pi_{\underline{\alpha}\overline{\beta}}$ and $a \leq \frac{1}{2}$ imply $H < 0$. Therefore

$$bU(w_{BS}) > \frac{\left[\varepsilon_A \Pr\left(\overline{\beta}\right) - \varepsilon_B \Pr\left(\underline{\alpha}\right)\right](1-b)\,U(w_{BF}) + \Pr\left(\overline{\beta}\right)(1-2\varepsilon_A)(1-a)\,U(w_{AF})}{G}; \quad (38)$$

$$bU(w_{BS}) < \frac{\left[(1-\varepsilon_B)\Pr\left(\underline{\alpha}\right) - \varepsilon_A \Pr\left(\underline{\beta}\right)\right](1-b)\,U(w_{BF}) - \Pr\left(\underline{\beta}\right)(1-2\varepsilon_A)(1-a)\,U(w_{AF})}{H} \quad (39)$$

A necessary condition for (38) and (39) is that the right hand side of (39) is greater than the right hand side of (38) and this implies $w_{BF} < w_{AF}$. From (36) we obtain

$$w_{BF} < w_{AF} < w_{AS} < w_{BS}.$$

Assume now $a > \frac{1}{2}$. From (36) we obtain that

$$(1-a)\,U(w_{AF}) = bU(w_{BS}) + (1-b)\,U(w_{BF}) - aU(w_{AS})\,. \quad (40)$$

Substituting (40) into (12) and (13)

$$(1-b)\,U\,(w_{BF})\,J <$$
$$\left[(1-\varepsilon_B)\Pr\left(\underline{\alpha}\right) - (1-\varepsilon_A)\Pr\left(\overline{\beta}\right)\right] bU\,(w_{BS}) + \Pr\left(\overline{\beta}\right)(1-2\varepsilon_A)\,aU\,(w_{AS})\,;$$

$$(1-b)\,U\,(w_{BF})\,K >$$
$$\left[(1-\varepsilon_B)\Pr\left(\overline{\alpha}\right) - \varepsilon_A\Pr\left(\overline{\beta}\right)\right] bU\,(w_{BS}) - \Pr\left(\overline{\beta}\right)(1-2\varepsilon_A)\,aU\,(w_{AS})\,;$$

where

$$\begin{aligned} J &= (1-\varepsilon_A)\Pr\left(\overline{\beta}\right) - \varepsilon_B\Pr\left(\underline{\alpha}\right); \\ K &= \varepsilon_A\Pr\left(\overline{\beta}\right) - \varepsilon_B\Pr\left(\overline{\alpha}\right). \end{aligned}$$

Simple algebra shows that J is strictly positive and that $\pi_{\overline{\alpha}\underline{\beta}} < \pi_{\underline{\alpha}\overline{\beta}}$ and $a > \frac{1}{2}$ imply $K > 0$. Therefore

$$(1-b)\,U\,(w_{BF}) < \frac{\left[(1-\varepsilon_B)\Pr\left(\underline{\alpha}\right) - (1-\varepsilon_A)\Pr\left(\overline{\beta}\right)\right] bU\,(w_{BS}) + \Pr\left(\overline{\beta}\right)(1-2\varepsilon_A)\,aU\,(w_{AS})}{J} \tag{41}$$
$$(1-b)\,U\,(w_{BF}) > \frac{\left[(1-\varepsilon_B)\Pr\left(\overline{\alpha}\right) - \varepsilon_A\Pr\left(\overline{\beta}\right)\right] bU\,(w_{BS}) - \Pr\left(\overline{\beta}\right)(1-2\varepsilon_A)\,aU\,(w_{AS})}{K}. \tag{42}$$

A necessary condition for (41) and (42) is that the right hand side of (41) is greater than the right hand side of (42) and this implies $w_{BS} < w_{AS}$. From (36) we obtain

$$w_{AF} < w_{BF} < w_{BS} < w_{AS}.$$

■

Proof of Proposition 1 Suppose that $w = (w_{AF}, w_{AS}, w_{BF}, w_{BS})$ solves (P) and $w_{AF} = w_{BF} = w_F$ and $w_S = w_{AS} = w_{BS}$. For (IC-$\alpha\beta$) to hold, it must be the case that $w_S \geq w_F$. Now, if $w_S = w_F$, (IC-A) and (IC-B) would not hold. Therefore, if w implements the efficient action profile $\hat{s}$, $w_S > w_F$ and (IC-$\alpha\beta$) hold with slack.

Assume first that $a > b$. Then, $aU(w_S)+(1-a)U(w_F) > bU(w_S)+(1-b)U(w_F)$, so if (IC-$A$) holds, then (IC-$B$) holds with slack. Consider now contract w', with $w'_{AF} = w_F$, $w'_{AS} = w_S - \Delta_A$, $w'_{BF} = w_F + \Delta_B$ and $w'_{BS} = w_S$, where Δ_A and Δ_B are such that $V(w) = V(w')$, so that

(PC) holds for w'. Further, for Δ_A and Δ_B small enough, (IC-B) and (IC-$\alpha\beta$) will also hold, since they held with slack for w. Finally, (IC-A) will hold as well, since $V(w') = V(w) \geq aU(w_S) + (1-a)U(w_F) > aU(w'_{AS}) + (1-a)U(w_F)$. Therefore, w' implements the efficient action profile $\hat{s}$. It only remains to be checked that it lowers the expected cost. Letting $h(x|s^*)$ denote the probability of public history x given the efficient action profile $\hat{s}$, it follows from $V(w) = V(w')$ that for Δ_A and Δ_B small enough, $h(BS|s^*)U'(w_F)\Delta_B = h(AS|s^*)U'(w_S)\Delta_A$, so

$$\Delta_B = \Delta_A \frac{h(AS|s^*)U'(w_S)}{h(BS|s^*)U'(w_F)} \tag{43}$$

Therefore, the change in the expected cost is

$$h(BS|s^*)\Delta_B - h(AS|s^*)\Delta_A = \Delta_A h(AS|s^*)\left(\frac{U'(w_S)}{U'(w_F)} - 1\right) < 0, \tag{44}$$

since risk aversion implies that $U'(w_S) < U'(w_F)$.

Assume now that $a = b$ and that $\varepsilon_A < \varepsilon_B$. This implies that the manager will base his decision only on signal α, the most precise signal. Consider now a change $dw = (0, 0, \Delta_F, \Delta_S)$, such that $\Delta_F > 0$, $\Delta_S < 0$ and $\mathbf{D}V dw = 0$ (i.e., the change keeps $V(w)$ unchanged). By definition of dw, (PC) and (IC-A) are unchanged. Further, since (IC-$\alpha\beta$) hold with slack, they will continue to hold for dw small enough. Now we check if (IC-B) still holds. By construction of dw

$$\pi_{\underline{\alpha}\overline{\beta}} \mathbf{D}V_{\overline{B}} dw + \pi_{\underline{\alpha}\underline{\beta}} \mathbf{D}V_{\underline{B}} dw = 0,$$

i.e,

$$\mathbf{D}V_{\underline{B}} dw = -\left(\frac{\pi_{\underline{\alpha}\overline{\beta}}}{\pi_{\underline{\alpha}\underline{\beta}}}\right) \mathbf{D}V_{\overline{B}} dw.$$

Consider now

$$\begin{aligned} \mathbf{D}V_B dw &= \Pr\left(\underline{\beta}\right) \mathbf{D}V_{\underline{B}} dw + \Pr\left(\overline{\beta}\right) \mathbf{D}V_{\overline{B}} dw = -\Pr\left(\underline{\beta}\right)\left(\frac{\pi_{\underline{\alpha}\overline{\beta}}}{\pi_{\underline{\alpha}\underline{\beta}}}\right) \mathbf{D}V_{\overline{B}} dw + \Pr\left(\overline{\beta}\right) \mathbf{D}V_{\overline{B}} dw \\ &= \mathbf{D}V_{\overline{B}} dw \left(\Pr\left(\overline{\beta}\right) - \Pr\left(\underline{\beta}\right)\left(\frac{\pi_{\underline{\alpha}\overline{\beta}}}{\pi_{\underline{\alpha}\underline{\beta}}}\right)\right) = 0, \end{aligned}$$

since

$$\frac{\pi_{\underline{\alpha}\overline{\beta}}}{\pi_{\underline{\alpha}\underline{\beta}}} = \frac{\Pr(\underline{\alpha})\Pr(\overline{\beta})}{\Pr(\underline{\alpha})\Pr(\underline{\beta})} = \frac{\Pr(\overline{\beta})}{\Pr(\underline{\beta})}.$$

Therefore, (IC-B) will hold as well. Finally, it only remains to check that dw decreases the expected cost of the contract. Letting $h(x|s^*)$ denote the probability of public history x given the efficient action profile $\hat{s}$, it follows from $V(w) = V(w')$ that for Δ_F and Δ_S small enough, $h(BS|s^*)U'(w_S)\Delta_S + h(BF|s^*)U'(w_F)\Delta_F = 0$, so

$$\Delta_S = -\Delta_F \frac{h(BF|s^*)U'(w_F)}{h(BS|s^*)U'(w_S)}.$$

Therefore, the change in the expected cost is

$$h(BS|s^*)\Delta_S + h(BF|s^*)\Delta_F = \Delta_F h(BF|s^*) \left(1 - \frac{U'(w_F)}{U'(w_S)}\right) < 0,$$

since risk aversion implies that $U'(w_S) < U'(w_F)$.

Finally, we show that $a = b$ and $\varepsilon_A = \varepsilon_B$ imply $w_{AS} = w_{BS} > w_{AF} = w_{BF}$. Notice first that $a = b$ and $\varepsilon_A = \varepsilon_B$ imply $\pi_{\overline{\alpha}\underline{\beta}} = \pi_{\underline{\alpha}\overline{\beta}}$. By Lemma 9 we know that (9) is binding. Assume first that (10) is binding. In Lemma 16 we have shown that (9) and (10) binding and $\pi_{\overline{\alpha}\underline{\beta}} = \pi_{\underline{\alpha}\overline{\beta}}$, imply that (13) and (14) are also binding. From (9) and (10) with equality we obtain

$$aU(w_{AS}) + (1-a)U(w_{AF}) = aU(w_{BS}) + (1-a)U(w_{BF}), \tag{45}$$

and from (14) with equality we obtain

$$\underline{a}U(w_{AS}) + (1-\underline{a})U(w_{AF}) = \underline{a}U(w_{BS}) + (1-\underline{a})U(w_{BF}). \tag{46}$$

Given that $a > \underline{a}$, (45) and (46) hold if and only if $U(w_{AS}) = U(w_{BS})$ and $U(w_{AF}) = U(w_{BF})$ which imply $w_{AS} = w_{BS}$ and $w_{AF} = w_{BF}$.

Assume now that (10) is not binding. From (10) not binding and (9) binding we obtain

$$aU(w_{AS}) + (1-a)U(w_{AF}) > aU(w_{BS}) + (1-a)U(w_{BF}). \tag{47}$$

By Lemma 10 we know that (14) is binding and this implies (46). Notice now that a necessary condition for (46) and (47) to both hold is that $U(w_{AS}) > U(w_{BS})$ which implies $w_{AS} > w_{BS}$. Recall now that in Lemma 17 we showed that if (10) and (14) are both binding, (13) cannot be binding. But Lemma 14 shows that this implies $U(w_{BS}) > U(w_{AS})$ and a contradiction arises.

■

Proof of Proposition 2

Notice that $\varepsilon_A = \varepsilon_B$ implies that $\pi_{\overline{\alpha}\underline{\beta}} > \pi_{\underline{\alpha}\overline{\beta}}$ and that $\underset{\sim}{\varepsilon}_A(\varepsilon_B) < \varepsilon_A < \widetilde{\varepsilon}_A(\varepsilon_B)$. The proof follows from Lemmas 13 and 14 ■

Proof of Proposition 3

The proof follows from Lemmas 13, 14, 15, 16, 17, 20, and 21.■

A.5 Proof of Proposition 4

We characterize the salary for part (1). Since in the following we never use the ex-ante efficiency of A ($a > b$), the proof for part (2) is symmetric.

If the efficient decision is sensitive to α, the optimal contract is a solution to:

$$\begin{aligned}
\min_{w \in \mathbb{R}^4} \quad & \pi_{\overline{\alpha}\underline{\beta}}[\overline{a}w_{AS} + (1-\overline{a})w_{AF}] + \pi_{\underline{\alpha}\overline{\beta}}[\overline{b}w_{BS} + (1-\overline{b})w_{BF}] + \\
& + \pi_{\overline{\alpha}\overline{\beta}}[\underline{a}w_{AS} + (1-\underline{a})w_{AF}] + \pi_{\underline{\alpha}\underline{\beta}}[\underline{b}w_{BS} + (1-\underline{b})w_{BF}] \\
s.t. \quad & \pi_{\overline{\alpha}\underline{\beta}}V_{\overline{A}} + \pi_{\underline{\alpha}\overline{\beta}}V_{\overline{B}} + \pi_{\overline{\alpha}\overline{\beta}}V_{\overline{A}} + \pi_{\underline{\alpha}\underline{\beta}}V_{\underline{B}} - g \geq \overline{U} & (48) \\
& \pi_{\overline{\alpha}\underline{\beta}}V_{\overline{A}} + \pi_{\underline{\alpha}\overline{\beta}}V_{\overline{B}} + \pi_{\overline{\alpha}\overline{\beta}}V_{\overline{A}} + \pi_{\underline{\alpha}\underline{\beta}}V_{\underline{B}} - g \geq \pi_{\overline{\alpha}\underline{\beta}}V_{\overline{A}} + \pi_{\underline{\alpha}\overline{\beta}}V_{\underline{A}} + \pi_{\overline{\alpha}\overline{\beta}}V_{\overline{A}} + \pi_{\underline{\alpha}\underline{\beta}}V_{\underline{A}} & (49) \\
& \pi_{\overline{\alpha}\underline{\beta}}V_{\overline{A}} + \pi_{\underline{\alpha}\overline{\beta}}V_{\overline{B}} + \pi_{\overline{\alpha}\overline{\beta}}V_{\overline{A}} + \pi_{\underline{\alpha}\underline{\beta}}V_{\underline{B}} - g \geq \pi_{\overline{\alpha}\underline{\beta}}V_{\underline{B}} + \pi_{\underline{\alpha}\overline{\beta}}V_{\overline{B}} + \pi_{\overline{\alpha}\overline{\beta}}V_{\overline{B}} + \pi_{\underline{\alpha}\underline{\beta}}V_{\underline{B}} & (50) \\
& V_{\overline{A}} \geq V_{\underline{B}} & (51) \\
& V_{\overline{B}} \geq V_{\underline{A}} & (52) \\
& V_{\overline{A}} \geq V_{\overline{B}} & (53) \\
& V_{\underline{B}} \geq V_{\underline{A}} & (54)
\end{aligned}$$

We characterize the optimal contract through a series of Lemmas.

LEMMA 22 *(1)* (51) *and* (53) *cannot be both binding. (2)* (52) *and* (54) *cannot be both binding.*

Proof. (1) Rewrite (50) as :

$$\pi_{\overline{\alpha}\underline{\beta}}(V_{\overline{A}} - V_{\underline{B}}) + \pi_{\overline{\alpha}\overline{\beta}}(V_{\overline{A}} - V_{\overline{B}}) \geq g > 0. \tag{55}$$

If (51) and (53) are both binding, the left hand side of (55) is equal to 0 and a contradiction is obtained.

(2) Rewrite (49) as

$$\pi_{\underline{\alpha}\overline{\beta}}(V_{\overline{B}} - V_{\underline{A}}) + \pi_{\underline{\alpha}\underline{\beta}}(V_{\underline{B}} - V_{\underline{A}}) \geq g > 0. \tag{56}$$

If (52) and (54) are both binding, the left hand side of (56) is equal to 0 and a contradiction is obtained. ■

LEMMA 23 $w_{AS} \geq w_{AF}$.

Proof. From (53) and (52), $V_{\overline{A}} \geq V_{\underline{A}}$. Since $\overline{a} > \underline{a}$, this requires $w_{AS} \geq w_{AF}$. ■

LEMMA 24 $w_{BS} \geq w_{BF}$.

Proof. Suppose that $w_{BS} < w_{BF}$. This implies $V_{\underline{B}} > V_{\overline{B}}$. Therefore, (51) implies that (53) is not binding and (52) implies that (54) is not binding.

Consider a change $dw = (0, 0, \Delta_{BF}, \Delta_{BS})$ such that $\Delta_{BF} < 0$, $\Delta_{BS} > 0$, and $\mathbf{D}V dw = 0$, i.e., the change keeps V unchanged[5]. Given that $\Delta_{BF} < 0$, $\Delta_{BS} > 0$, this requires

$$\mathbf{D}V_{\overline{B}} dw > 0 \tag{57}$$

$$\mathbf{D}V_{\underline{B}} dw < 0. \tag{58}$$

In the following, we will show that there exist Δ_{BF} and Δ_{BS} sufficiently small in absolute value that (i) satisfy constraints (48)-(54); (ii) the expected cost of the contract is lower.

(i) By definition of dw, (48) and (49) are unchanged. From (57) and (58), (51) and (52) also hold. Moreover, since (53) and (54) are not binding, they also hold for Δ_{BF} and Δ_{BS} sufficiently small in absolute value.

Notice now that by construction of dw:

$$\pi_{\underline{\alpha}\overline{\beta}}\mathbf{D}V_{\overline{B}} dw + \pi_{\underline{\alpha}\underline{\beta}}\mathbf{D}V_{\underline{B}} dw = 0, \tag{59}$$

and a sufficient condition for (50) to hold is that

$$\pi_{\overline{\alpha}\underline{\beta}}\mathbf{D}V_{\underline{B}} dw + \pi_{\overline{\alpha}\overline{\beta}}\mathbf{D}V_{\overline{B}} dw = 0 \tag{60}$$

[5] $\mathbf{D}$ denotes the Jacobian operator.

because this implies that the right hand side of (50) is unchanged (the left hand side is unchanged by construction of dw). From (59)

$$\mathbf{D}V_{\underline{B}}dw = -\frac{\pi_{\underline{\alpha}\overline{\beta}}}{\pi_{\underline{\alpha}\underline{\beta}}}\mathbf{D}V_{\overline{B}}dw. \tag{61}$$

Substituting (61) into (60) we obtain

$$\begin{aligned} -\pi_{\overline{\alpha}\underline{\beta}}\frac{\pi_{\underline{\alpha}\overline{\beta}}}{\pi_{\underline{\alpha}\underline{\beta}}}\mathbf{D}V_{\overline{B}}dw + \pi_{\overline{\alpha}\overline{\beta}}\mathbf{D}V_{\overline{B}}dw &= \mathbf{D}V_{\overline{B}}dw\left(\pi_{\overline{\alpha}\overline{\beta}} - \pi_{\overline{\alpha}\underline{\beta}}\frac{\pi_{\underline{\alpha}\overline{\beta}}}{\pi_{\underline{\alpha}\underline{\beta}}}\right) \\ &= \pi_{\overline{\alpha}\underline{\beta}}\mathbf{D}V_{\overline{B}}dw\left(\frac{\pi_{\overline{\alpha}\overline{\beta}}}{\pi_{\overline{\alpha}\underline{\beta}}} - \frac{\pi_{\underline{\alpha}\overline{\beta}}}{\pi_{\underline{\alpha}\underline{\beta}}}\right) = 0 \end{aligned} \tag{62}$$

with (62) following from the independence of A and B and of α and β that imply that

$$\frac{\pi_{\overline{\alpha}\overline{\beta}}}{\pi_{\overline{\alpha}\underline{\beta}}} = \frac{\Pr\left(\overline{\alpha},\overline{\beta}\right)}{\Pr\left(\overline{\alpha},\underline{\beta}\right)} = \frac{\Pr\left(\overline{\beta}\right)}{\Pr\left(\underline{\beta}\right)} = \frac{\Pr\left(\underline{\alpha},\overline{\beta}\right)}{\Pr\left(\underline{\alpha},\underline{\beta}\right)} = \frac{\pi_{\underline{\alpha}\overline{\beta}}}{\pi_{\underline{\alpha}\underline{\beta}}}.$$

(ii) Given that dw is such that $\mathbf{D}V dw = 0$

$$\left(\pi_{\underline{\alpha}\overline{\beta}}\overline{b} + \pi_{\underline{\alpha}\underline{\beta}}\underline{b}\right)U'\left(w_{BS}\right)\Delta_{BS} + \left(\pi_{\underline{\alpha}\overline{\beta}}\left(1-\overline{b}\right) + \pi_{\underline{\alpha}\underline{\beta}}\left(1-\underline{b}\right)\right)U'\left(w_{BF}\right)\Delta_{BF} = 0$$

which implies that

$$\Delta_{BS} = -\frac{\left(\pi_{\underline{\alpha}\overline{\beta}}\left(1-\overline{b}\right) + \pi_{\underline{\alpha}\underline{\beta}}\left(1-\underline{b}\right)\right)U'\left(w_{BF}\right)\Delta_{BF}}{\left(\pi_{\underline{\alpha}\overline{\beta}}\overline{b} + \pi_{\underline{\alpha}\underline{\beta}}\underline{b}\right)U'\left(w_{BS}\right)} \tag{63}$$

The change in the expected salary is

$$\left(\pi_{\underline{\alpha}\overline{\beta}}\overline{b} + \pi_{\underline{\alpha}\underline{\beta}}\underline{b}\right)\Delta_{BS} + \left(\pi_{\underline{\alpha}\overline{\beta}}\left(1-\overline{b}\right) + \pi_{\underline{\alpha}\underline{\beta}}\left(1-\underline{b}\right)\right)\Delta_{BF}. \tag{64}$$

Substituting (63) into (64) shows that the change in the expected salary is negative. ■

LEMMA 25 $w_{BF} \geq w_{BS}$.

Proof. The proof is symmetric to the proof of Lemma 24 with the only difference arising from the fact that in this case (53) implies that (51) is not binding and (54) implies that (52) is not binding. ■

LEMMA 26 $w_{AS} > w_{BS} = w_{BF} > w_{AF}$.

Proof. $w_{BS} = w_{BF}$ follows from Lemmas 24 and 25. From Lemma 22, $V_{\overline{A}} > V_{\underline{B}} = V_{\overline{B}}$, since (51) and (53) cannot be both binding. Therefore, (51) and (54) imply $V_{\overline{A}} > V_{\underline{B}} \geq V_{\underline{A}}$, which implies $V_{\overline{A}} > V_{\underline{A}}$, which implies $w_{AS} > w_{AF}$. Finally, (51) to (54) imply that $w_{AS} > w_{BS} = w_{BF} > w_{AF}$. ■

Proof of Proposition 4

Immediate from Lemma 26.

■

A.6 Proof of Proposition 5

The optimal contract is a solution to

$$\min_{w \in \mathbb{R}^4} \quad \pi_{\overline{\alpha}}[\overline{a} w_{AS} + (1-\overline{a}) w_{AF}] + \pi_{\underline{\alpha}}[\overline{b} w_{BS} + (1-\overline{b}) w_{BF}] \tag{65}$$

$$\text{s.t.} \quad \pi_{\overline{\alpha}} V_{\overline{A}} + \pi_{\underline{\alpha}} V_{\overline{B}} - g \geq \overline{U} \tag{66}$$

$$\pi_{\overline{\alpha}} V_{\overline{A}} + \pi_{\underline{\alpha}} V_{\overline{B}} - g \geq \pi_{\overline{\alpha}} V_{\overline{A}} + \pi_{\underline{\alpha}} V_{\underline{A}} \tag{67}$$

$$\pi_{\overline{\alpha}} V_{\overline{A}} + \pi_{\underline{\alpha}} V_{\overline{B}} - g \geq \pi_{\overline{\alpha}} V_{\underline{B}} + \pi_{\underline{\alpha}} V_{\overline{B}} \tag{68}$$

$$V_{\overline{A}} \geq V_{\underline{B}} \tag{69}$$

$$V_{\overline{B}} \geq V_{\underline{A}} \tag{70}$$

Because of the assumptions on the utility function, the solution to this problem is interior and satisfies the following first order conditions:

$$\frac{1}{U'(w_{AS})} = \frac{(\lambda_P + \lambda_B + \lambda_1)(1-\varepsilon) - (\lambda_A + \lambda_2)\varepsilon}{(1-\varepsilon)} \tag{71}$$

$$\frac{1}{U'(w_{AF})} = \frac{(\lambda_P + \lambda_B + \lambda_1)\varepsilon - (\lambda_A + \lambda_2)(1-\varepsilon)}{\varepsilon} \tag{72}$$

$$\frac{1}{U'(w_{BS})} = \frac{(\lambda_P + \lambda_A + \lambda_2)(1-\varepsilon) - (\lambda_B + \lambda_1)\varepsilon}{(1-\varepsilon)} \tag{73}$$

$$\frac{1}{U'(w_{BF})} = \frac{(\lambda_P + \lambda_A + \lambda_2)\varepsilon - (\lambda_B + \lambda_1)(1-\varepsilon)}{\varepsilon} \tag{74}$$

We prove the Proposition through a sequence of Lemmas.

LEMMA 27 (67) *and* (70) *cannot hold with slack simultaneously.*

Proof. Suppose that (67) and (70) both hold with slack. From (71)-(74) we obtain

$$w_{AF} = w_{AS} = w_{BF} = w_{BS}$$

and a contradiction to (67) and (68) arises. ■

LEMMA 28 (69) *and* (70) *hold with slack.*

Proof. If (69) is binding, (68) is violated. If (70) is binding, (67) is violated. ■

LEMMA 29 (67) *is binding.*

Proof. Immediate from Lemmas 27 and 28 ■

LEMMA 30 (68) *is binding.*

Proof. Suppose that (68) holds with slack. From (71)-(74) we obtain

$$w_{AF} \leq w_{AS} \leq w_{BF} = w_{BS}$$

which violates (69). ■

LEMMA 31 $U(w_{AS}) - U(w_{BF}) < U(w_{BS}) - U(w_{AF})$.

Proof. From Lemmas 29 and 30 we have

$$aU(w_{AS}) + (1-a)U(w_{AF}) = (1-a)U(w_{BS}) + aU(w_{BF}),$$

which is equivalent to

$$a[U(w_{AS}) - U(w_{BF})] = (1-a)[U(w_{BS}) - U(w_{AF})]. \tag{75}$$

A necessary condition for (75) to hold is

$$U(w_{AS}) - U(w_{BF}) < U(w_{BS}) - U(w_{AF}).$$

■

LEMMA 32 $w_{AS} > w_{AF}$ *and* $w_{BS} > w_{BF}$.

Proof. Immediate from Lemmas 28, 29 and 30 and from (71)-(74). ■

LEMMA 33 $w_{AF} < w_{BF} \leq w_{AS} < w_{BS}$.

Proof. From Lemmas 28, 29 and 30 and from (71)-(74) we obtain that

$$w_{BS} \geq w_{AS} \iff w_{BF} \geq w_{AF} \tag{76}$$

$$w_{BS} \leq w_{AS} \iff w_{BF} \leq w_{AF} \tag{77}$$

Suppose that $w_{BS} = w_{AS}$. From (76) and (77) this holds if and only if $w_{BF} = w_{AF}$. This in turn implies that (67) and (68) cannot be both binding, a contradiction to either of Lemmas 29 and 30.

Suppose now that $w_{BS} < w_{AS}$. From (76) and (77) this holds if and only if $w_{BF} < w_{AF}$. By Lemma 32, this implies one of the following

$$w_{BF} < w_{BS} \leq w_{AF} < w_{AS} \tag{78}$$

$$w_{BF} < w_{AF} < w_{BS} < w_{AS} \tag{79}$$

Notice that (78) contradicts Lemma 31 and (79) implies that (67) and (68) cannot both be binding, therefore contradicting at least one of Lemmas 29 and 30.

Suppose therefore that $w_{BS} > w_{AS}$. From (76) and (77) this holds if and only if $w_{BF} > w_{AF}$. By Lemma 32, this implies one of the following

$$w_{AF} < w_{AS} < w_{BF} < w_{BS} \tag{80}$$

$$w_{AF} < w_{BF} \leq w_{AS} < w_{BS} \tag{81}$$

Because (80) violates (69) the only possible ranking is the one in (81) and the result follows. ■

Proof of Proposition 5

Immediate from Lemma 33.