

Online Appendix for “Tractability in Incentive Contracting”

Alex Edmans and Xavier Gabaix

November 9, 2009

D Multidimensional Signal and Action

While the core model involves a single signal and action, this section shows that our contract is robust to a setting of multidimensional signals and actions. For brevity, we only analyze the discrete-time one-period case, since the continuous time extension is similar. The agent now takes a multidimensional action $\mathbf{a} \in \mathcal{A}$, which is a compact subset of \mathbb{R}^I for some integer I . (Note that in this section, bold font has a different usage than in the proof of Theorem 1.) The signal is also multidimensional:

$$\mathbf{r} = \mathbf{b}(\mathbf{a}) + \boldsymbol{\eta},$$

where $\boldsymbol{\eta}, \mathbf{r} \in \mathbb{R}^S$, and $\mathbf{b}: \mathcal{A} \in \mathbb{R}^I \rightarrow \mathbb{R}^S$. The signal and action can be of different dimensions. In the core model, $S = I = 1$ and $\mathbf{b}(\mathbf{a}) = a$. As before, the contract is $c(\mathbf{r})$ and the indirect felicity function is $V(\mathbf{r}) = v(c(\mathbf{r}))$. The following Proposition states the optimal contract.

Proposition 5 (*Optimal contract, discrete time, multidimensional signal and action*). Define the $I \times S$ matrix $L = \mathbf{b}'(\mathbf{a}^*)^\top$ i.e. explicitly $L_{ij} = \frac{\partial \mathbf{b}_j}{\partial a_i}(a_1^*, \dots, a_I^*)$, and assume that there is a vector $\theta \in \mathbb{R}^S$ such that

$$L\theta = g'(\mathbf{a}^*), \tag{63}$$

i.e., explicitly:

$$\forall i = 1 \dots I, \sum_{j=1}^S \frac{\partial \mathbf{b}_j}{\partial a_i}(a_1^*, \dots, a_I^*) \theta_j = \frac{\partial g}{\partial a_i}(a_1^*, \dots, a_I^*).$$

The following contract is optimal. The agent is paid

$$c(\mathbf{r}) = v^{-1}(\theta \mathbf{r} + K(\mathbf{r})), \tag{64}$$

i.e., explicitly, $c(\mathbf{r}) = v^{-1}\left(\sum_{j=1}^S \theta_j r_j + K(r_1, \dots, r_n)\right)$, where the function $K(\cdot)$ is the solution of the following optimization problem:

$$\min_{K(\cdot)} \mathbb{E}[K(\mathbf{b}(\mathbf{a}^*) + \boldsymbol{\eta})] \text{ subject to}$$

$$\forall \mathbf{r}, LK'(\mathbf{r}) = 0 \tag{65}$$

$$\mathbb{E}[u(\theta(\mathbf{b}(\mathbf{a}^*) + \boldsymbol{\eta}) + K(\mathbf{b}(\mathbf{a}^*) + \boldsymbol{\eta}) - g(\mathbf{a}^*))] \geq \underline{u}.$$

Proof. Here we derive the first-order condition; the remainder of the proof is as in Theorem 1 of the main paper. Incentive compatibility requires that, for all $\boldsymbol{\eta}$

$$\mathbf{a}^* \in \arg \max_{\mathbf{a}} V(\mathbf{b}(\mathbf{a}) + \boldsymbol{\eta}) - g(\mathbf{a}),$$

and so:

$$V'(\mathbf{b}(\mathbf{a}^*) + \boldsymbol{\eta}) \mathbf{b}'(\mathbf{a}^*) - g'(\mathbf{a}^*) = 0, \quad (66)$$

where V' is a S -dimensional vector, $\mathbf{b}'(\mathbf{a}^*)$ is a $S \times I$ matrix, and $g'(\mathbf{a}^*)$ is a I -dimensional vector. Integrating (66) gives: $V(\mathbf{r}) = \boldsymbol{\theta}\mathbf{r} + K(\mathbf{r})$, where $\boldsymbol{\theta}\mathbf{r} = \sum_{i=1}^S \theta_i r_i$, and $LK'(\mathbf{r}) = 0$.

Note that $K(\mathbf{r})$ is now a function and so determined by solving an optimization problem. In the core model, K is a constant and determined by solving an equality. ■

We now analyze two specific applications of this extension.

Two signals. The agent takes a single action, but there are two signals of performance:

$$r_1 = a + \varepsilon_1, \quad r_2 = a + \varepsilon_2.$$

In this case, $L = (1 \ 1)$. Therefore, with $\boldsymbol{\theta} = (\theta_1, \theta_2) \in \mathbb{R}^2$, (63) becomes: $\theta_1 + \theta_2 = g'(a^*)$. For example, we can take $\theta_1 = \theta_2 = g'(a^*)/2$. Next, (65) becomes: $\partial K/\partial r_1 + \partial K/\partial r_2 = 0$. It is well known that this can be integrated into: $K(r_1, r_2) = k(r_1 - r_2)$ for a function k . Hence, the optimal contract can be written:

$$c = v^{-1} \left(g'(a^*) \left(\frac{r_1 + r_2}{2} \right) + k(r_1 - r_2) \right),$$

where the function $k(\cdot)$ is chosen to minimize the cost of the contract subject to the participation constraint. As in Holmstrom (1979), all informative signals should be used to determine the agent's compensation.

Relative performance evaluation. Again, there is a single action and two signals, but the second signal is independent of the agent's action, as in Holmstrom (1982):

$$r_1 = a + \varepsilon_1, \quad r_2 = \varepsilon_2$$

In this case, $L = (1 \ 0)$. Therefore, with $\boldsymbol{\theta} = (\theta_1, \theta_2) \in \mathbb{R}^2$, (63) becomes: $\theta_1 = g'(a^*)$. Next, (65) becomes: $\partial K/\partial r_1 = 0$, so that $K(r_1, r_2) = k(r_2)$ for a function k . Hence, the optimal contract can be written:

$$c = v^{-1} (g'(a^*) r_1 + k(r_2)).$$

The second signal enters the contract even though it is unaffected by the agent's action, since it may be correlated with the noise in the first signal.

E Extension to The Optimal Effort Level

E.1 Illustrations for Proposition 2

E.1.1 Affine Cost of Effort

While Theorem 3 shows that $A(\eta) = \bar{a}$ is optimal when Proposition 3 is satisfied, we now show that $A(\eta)$ can be exactly derived even if Theorem 3 does not hold and the maximum effort principle does not apply, if the cost function is linear – i.e. $g(a) = \theta a$, where $\theta > 0$.²⁴ We use the benefit function $b(a, \eta) = Sb_*(a, \eta)$ as in Section 3.2.

Proposition 6 (*Optimal contract with linear cost of effort*). *Let $g(a) = \theta a$, where $\theta > 0$. The following contract is optimal:*

$$c = v^{-1}(\theta r + K), \quad (67)$$

where K is a constant that makes the participation constraint bind ($\mathbb{E}[u(\theta\eta + K)] = \underline{u}$). For each η , the optimal effort $A(\eta)$ is determined by the following pointwise maximization:

$$A(\eta) \in \arg \max_{a \leq \bar{a}} Sb_*(a, \eta) - v^{-1}(\theta(a + \eta) + K). \quad (68)$$

When the agent is indifferent between an action a and $A(\eta)$, we assume that he chooses action $A(\eta)$.

Proof. From Proposition 2, if the agent announces η , he should receive a felicity of $V(\eta) = g(A(\eta)) + \int_{\eta_*}^{\eta} \theta dx + K = \theta(A(\eta) + \eta) + K$. Since $r = A(\eta) + \eta$ on the equilibrium path, a contract $c = v^{-1}(\theta r + K)$ will implement $A(\eta)$. To find the optimal action, the principal's problem is:

$$\max_{A(\eta)} \mathbb{E} [Sb_*(\min(A(\eta), \bar{a}), \eta)] - \mathbb{E} [v^{-1}(\theta(A(\eta) + \eta) + K)]$$

which is solved by pointwise maximization, as in (68). ■

The main advantage of the above contract is that it can be exactly solved regardless of S and so it is applicable even for small firms (or rank-and-file employees who affect a small output). For instance, consider a benefit function $b_*(a, \eta) = b_0 + ae^\eta$, where $b_0 > 0$, so that the marginal productivity of effort is increasing in the noise, and utility function $u(\ln c - \theta a)$ with $\theta \in (0, 1)$. Then, the solution of (68) is:

$$A(\eta) = \min \left(\frac{1 - \theta}{\theta} \eta + \frac{1}{\theta} (\ln S - K - \ln \theta), \bar{a} \right).$$

The optimal effort level increases linearly with the noise, until it reaches \bar{a} . The effort level is also weakly increasing in firm size.

²⁴Note that the linearity of $g(a)$ is still compatible with $u(v(c) - g(a))$ being strictly concave in (c, a) . Also, by a simple change of notation, the results extend to an affine rather than linear $g(a)$.

Note that, with a linear rather than strictly convex cost function, the agent is indifferent between all actions. His decision problem is $\max_a v(c(r)) - g(a)$, i.e. $\max_a \theta(\eta + a) + K - \theta a$, which is independent of a and thus has a continuum of solutions. As in, e.g., Grossman and Hart (1983), Proposition 6 therefore assumes that indeterminacies are resolved by the agent following the principal's recommended action, $A(\eta)$.

E.1.2 Exponential u and Linear v

We continue to assume that the maximum effort principle does not apply, and now consider the case where consider the HM assumptions of exponential utility and a pecuniary cost of effort, but do not impose Gaussian noise nor continuous time. We show that, as in HM, the same action function $A_t(\eta_t)$ is optimal in each period t . However, unlike in HM, $A_t(\eta_t)$ is not a constant independent of η_t . The intuition is that, if noise is low, the optimal contract may wish to reduce the required effort level to cushion the effect of low noise on the agent's utility.

Proposition 7 (*Constant target action, exponential utility and pecuniary cost of effort*). *Suppose the agent has a CARA utility function $u(x) = -e^{-\gamma x}$ and a linear felicity function $v(x) = x$, and suppose the benefit of effort in each period is a weakly concave function $b(a)$. Then, the optimal contract prescribes the same (possibly noise-dependent) action $A(\eta)$ in each period.*

Proof. Take an optimal contract specifying actions $A_1(\eta_1), \dots, A_T(\eta_1, \dots, \eta_T)$, and compensation $C(\eta_1, \dots, \eta_T)$. Start with period $t = T$. The optimality of the contract implies that for all $(\eta_1, \dots, \eta_{T-1})$, the choice of target action and compensation solve the optimization problem

$$\begin{aligned} & \max E_{\eta_T} [b(A_T(\eta_1, \dots, \eta_{T-1}, \eta_T)) - C(\eta_1, \dots, \eta_{T-1}, \eta_T)] \\ & \text{s.t. } \eta_T \in \arg \max_{\hat{\eta}} [-e^{-\gamma\{C(\eta_1, \dots, \eta_{T-1}, \hat{\eta}_T) - g(A(\eta_1, \dots, \eta_{T-1}, \hat{\eta}_T) + \hat{\eta}_T - \eta_T)\}}] , \\ & E_{\eta_T} [-e^{-\gamma\{C(\eta_1, \dots, \eta_{T-1}, \eta_T) - g(A(\eta_1, \dots, \eta_{T-1}, \eta_T))\}}] = \underline{u}(\eta_1, \dots, \eta_{T-1}). \end{aligned}$$

By Proposition 2, the cost of compensation for a given action $A_T(\eta_1, \dots, \eta_T)$ is minimized by

$$C(\eta_1, \dots, \eta_T) = g(A(\eta_1, \dots, \eta_T)) + \int_{\eta_*}^{\eta_T} A_T(\eta_1, \dots, \eta_{T-1}, x) dx + K(\eta_1, \dots, \eta_{T-1}),$$

so the principal solves a collection of problems

$$\max_{A(\cdot), K} E_{\eta_T} \left[b(A(\eta_T)) - g(A(\eta_T)) - \int_{\eta_*}^{\eta_T} A(x) dx - K \right] \quad (69)$$

$$\text{s.t. } E_{\eta_T} \left[-e^{-\gamma \int_{\eta_*}^{\eta_T} A(x) dx - \gamma K} \right] = \underline{u}(\eta_1, \dots, \eta^{T-1}) \quad (70)$$

for (possibly) varying $\underline{u}(\eta_1, \dots, \eta_{T-1})$. By concavity, the solutions of these problems for each $\underline{u}(\eta_1, \dots, \eta_{T-1})$ are unique. Moreover, this uniqueness implies that the solutions for different

values of $\underline{u}(\eta_1, \dots, \eta_{T-1})$ may differ only in the constant K . Therefore, the optimal target action $A_T(\eta_1, \dots, \eta_{T-1}, \eta_T)$ does not depend on $\eta_1, \dots, \eta_{T-1}$.

Now, since A_{T-1} is the only action that can depend on η_{T-1} , the above argument can be repeated for $t = T - 1, \dots, 1$. Hence, the optimal profile of actions $A_1(\eta_1), \dots, A_T(\eta_1, \dots, \eta_T)$ consists of repeating the same target action $A(\cdot)$, which is the unique solution of the problem (69)–(70). ■

Example A. Suppose $b(x) = Bx$, $g(x) = \frac{1}{2}Gx^2$, $\eta \sim U[\underline{\eta}, \bar{\eta}]$. Let $y(\eta) = \int_{\eta_*}^{\eta} a(x) dx + K$. Then, the optimal target action is the solution of

$$\begin{aligned} & \max_{a(\cdot), y(\cdot)} \int_{\underline{\eta}}^{\bar{\eta}} \left(Ba(x) - \frac{1}{2}Ga(x)^2 - y(x) \right) dx \\ & \text{s.t. } \int_{\underline{\eta}}^{\bar{\eta}} (-e^{-\gamma y(x)}) = \underline{u}, \\ & y'(x) = a(x). \end{aligned}$$

The Lagrangian of this problem is

$$\begin{aligned} \mathcal{L} &= \int_{\underline{\eta}}^{\bar{\eta}} \left(Ba(x) - y(x) - \frac{1}{2}Ga(x)^2 - \lambda e^{-\gamma y(x)} + \mu(x)(a(x) - y'(x)) \right) dx \\ &= \int_{\underline{\eta}}^{\bar{\eta}} \left(Ba(x) - y(x) - \frac{1}{2}Ga(x)^2 - \lambda e^{-\gamma y(x)} + \mu(x)a(x) + \mu'(x)y(x) \right) dx \\ &\quad - \mu(\bar{\eta})y(\bar{\eta}) + \mu(\underline{\eta})y(\underline{\eta}), \end{aligned}$$

where λ is the multiplier attached to the reservation utility constraint, and $\mu(x)$ is the multiplier for the equation linking $y(x)$ and $a(x)$. Note that \mathcal{L} is concave in $a(x)$ and $y(x)$. The first-order conditions are

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial a(x)} : & \quad B - Ga(x) + \mu(x) = 0, \\ \frac{\partial \mathcal{L}}{\partial y(x)} : & \quad -1 + \lambda \gamma e^{-\gamma y(x)} + \mu'(x) = 0, \\ \frac{\partial \mathcal{L}}{\partial y(\underline{\eta})}, \frac{\partial \mathcal{L}}{\partial y(\bar{\eta})} : & \quad \mu(\underline{\eta}) = \mu(\bar{\eta}) = 0. \end{aligned}$$

Substituting the first equality into the second we get

$$-1 + \lambda \gamma e^{-\gamma y(x)} + Ga'(x) = 0.$$

Rearranging and taking a logarithm gives

$$\ln(\lambda \gamma) - \gamma y(x) = \ln(1 - Ga'(x)).$$

Differentiating the last equality gives

$$-\gamma y'(x) = -G \frac{a''(x)}{1 - Ga'(x)},$$

which can be simplified into

$$a''(x) = \gamma a(x) (1 - Ga'(x)) / G.$$

So, the optimal action satisfies a second-order ODE with the boundary conditions

$$a(\underline{\eta}) = a(\bar{\eta}) = B/G,$$

and indeed does not depend on the reservation utility \underline{u} .

Example B. Take the same functions, $b(x) = Bx$, $g(x) = \frac{1}{2}Gx^2$ and suppose that the noise is Gaussian, $\eta \sim N(0, \sigma^2)$. We will be solving the optimization problem on the interval $[-z, z]$, and then take the limit as $z \rightarrow \infty$. Similar to Example A, the Lagrangian of the problem is

$$\begin{aligned} \mathcal{L} = \int_{-z}^z & \left(\left(Ba(x) - y(x) - \frac{1}{2}Ga(x)^2 - \lambda e^{-\gamma y(x)} \right) \phi\left(\frac{x}{\sigma}\right) + \mu(x)a(x) + \mu'(x)y(x) \right) dx \\ & - \mu(z)y(\bar{\eta}) + \mu(-z)y(\underline{\eta}), \end{aligned}$$

and the first-order conditions are

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial a(x)} : & \quad (B - Ga(x)) \phi\left(\frac{x}{\sigma}\right) + \mu(x) = 0, \\ \frac{\partial \mathcal{L}}{\partial y(x)} : & \quad (-1 + \lambda \gamma e^{-\gamma y(x)}) \phi\left(\frac{x}{\sigma}\right) + \mu'(x) = 0, \\ \frac{\partial \mathcal{L}}{\partial y(-z)}, \frac{\partial \mathcal{L}}{\partial y(z)} : & \quad \mu(-z) = \mu(z) = 0. \end{aligned}$$

Substituting the first equality into the second to eliminate $\mu(x)$, and taking note that

$$\frac{1}{\phi(x/\sigma)} \frac{d}{dx} (\phi(x/\sigma)) = -\frac{x}{\sigma^2},$$

we obtain

$$-1 + \lambda \gamma e^{-\gamma y(x)} + Ga'(x) - \frac{1}{\sigma^2} x (Ga(x) - B) = 0.$$

Rearranging and taking a logarithm gives

$$\ln(\lambda \gamma) - \gamma y(x) = \ln \left(1 - Ga'(x) + \frac{1}{\sigma^2} x (Ga(x) - B) \right).$$

Differentiating, taking note that $y'(x) = a(x)$, and rearranging yields the following: the optimal action is the limit as $z \rightarrow \infty$ of the solutions of

$$\begin{cases} a''(x) = \gamma a(x) \left[\frac{1}{G} - a'(x) + \frac{x}{\sigma^2} (a(x) - \frac{B}{G}) \right] + \frac{x}{\sigma^2} a'(x) + \frac{1}{\sigma^2} (a(x) - \frac{B}{G}) \\ a(-z) = a(z) = B/G. \end{cases}$$

E.2 Conditions for Maximum Effort Principle

Section 3.2 showed that the condition in Theorem 3,

$$\forall \eta, \forall a \leq \bar{a}, \partial_1 b(a(\eta), \eta) f(\eta) \geq \lambda(\bar{a}, \eta)$$

required for the maximum effort principle to hold, is satisfied if firm size S is sufficiently large. This extension considers other cases in which the above condition is satisfied, and shows sufficient conditions for the function $\lambda(\bar{a}, \eta)$.

By Proposition 2, the optimal contract is:

$$c(\eta) = v^{-1}(g(a(\eta)) + L(\eta) + K),$$

where $L(\eta) = \int_{\eta_*}^{\eta} g'(a(x)) dx$, η_* is an arbitrary constant in the support of η . The contract's cost is:

$$C[A] = E[v^{-1}(g(a(\eta)) + L(\eta) + K)].$$

Then we can take $\lambda(\bar{a}, \eta) = \max(0, \partial C[A] / \partial a(\eta))$, where $\partial C[A] / \partial a(\eta)$ is given by the following expression.²⁵

Proposition 8 *Assume that $\sup_{\eta} f(\eta) < \infty$. For an effort profile $a(\eta) + \eta$ satisfying the conditions of Proposition 2, the marginal cost of implementing effort $a(\eta)$ is:*

$$\begin{aligned} \frac{\partial C[A]}{\partial a(\eta)} &= \frac{g'(a(\eta))}{v'(c(\eta))} f(\eta) + \\ &g''(a(\eta)) \left\{ E \left[\frac{1}{v'(c(\tilde{\eta}))} 1_{\tilde{\eta} > \eta} \right] - E \left[\frac{1}{v'(c(\tilde{\eta}))} \right] \frac{E[u'(L(\tilde{\eta}) + K) 1_{\tilde{\eta} > \eta}]}{E[u'(L(\tilde{\eta}) + K)]} \right\}. \end{aligned} \quad (71)$$

where the expectation is taken over $\tilde{\eta}$.

The first term in (71), $\frac{g'(a(\eta))}{v'(c(\eta))} f(\eta)$, is the “local” compensating differential for inducing greater effort. Indeed, consider making the agent work δa more at point $\tilde{\eta}$. Let δc denote the

²⁵The proof is thus. Note that K satisfies $u = E[u(L(\eta) + K)]$. For simplicity, we assume $\eta_* < \eta$ (otherwise, we can just consider a lower η_*). Using $\partial L(\eta') / \partial a(\eta) = 1_{\eta' > \eta} g''(a(\eta))$, we have:

$$\frac{\partial K}{\partial a(\eta)} = \frac{-E[u'(L(\eta') + K) 1_{\eta' > \eta}]}{E[u'(L(\eta') + K)]} g''(a(\eta))$$

which implies (71).

additional pay that compensates him purely for the disutility of effort. We require

$$v(c(\eta)) - g(a) = v(c(\eta) + \delta c) - g(a + \delta a)$$

and so the additional pay is:

$$\delta c = \frac{g'(a)}{v'(c(\eta))} \delta a.$$

The $f(\eta)$ term in (71) simply multiplies it by the probability of observing noise η . The second term is the effect of a local change on the whole pattern of incentives: if $a(\eta)$ changes, it will affect the payment for the other noises $\eta' \neq \eta$, as indicated in Proposition 2. This change in the entire contract increases the agent's risk. Hence, the two terms capture the standard effects of implementing a greater effort level: direct disutility, plus inefficient risk-sharing caused by the sharper incentives required. The second term can be evaluated directly for concrete distributions; in addition, we can establish bounds on it to help verify whether Proposition 3 is satisfied. For instance, where noise has a finite upper bound $\bar{\eta}$, we obtain the following bound:

$$\frac{\partial C[A]}{\partial a(\eta)} \leq \frac{g'(a(\eta))}{v'(c(\eta))} f(\eta) + \frac{g''(a(\eta))}{v'(c(\bar{\eta}))} P(\tilde{\eta} > \eta).$$

Second, the upper bound for $\frac{\partial C[A]}{\partial a(\eta)}$ and thus $\lambda(\bar{a}, \eta)$ is simpler when noise is bounded both above and below. If $\text{supp } \eta = [\underline{\eta}, \bar{\eta}]$ and $g'''(x) \geq 0$ for all x . Then

$$\frac{\partial C[A]}{\partial a(\eta)} \leq \Lambda(\bar{a}, \eta) \equiv \frac{g'(\bar{a})f(\eta) + g''(\bar{a})\bar{F}(\eta)}{v'(v^{-1}(u^{-1}(\underline{u}) + g(\bar{a}) + (\bar{\eta} - \underline{\eta})g'(\bar{a})))}. \quad (72)$$

In particular, in (27), the function λ can be replaced by the function Λ . We observe that $\Lambda(\bar{a}, \eta)$ is increasing in \bar{a} .

The proof of (72) is thus. We observe that

$$L(\eta) + K \leq u^{-1}(\underline{u}) + (\bar{\eta} - \underline{\eta})g'(\bar{a}),$$

for any η . If it does not hold for some η_0 , then

$$L(\eta) + K = \int_{\eta_*}^{\eta} g'(a(x)) dx + K = L(\eta_0) + K + \int_{\eta_0}^{\eta} g'(a(x)) dx \geq L(\eta_0) + K - (\bar{\eta} - \underline{\eta})g'(\bar{a}) > u^{-1}(\underline{u})$$

for all η , and the constraint $E[u(L(\eta) + K)] = \underline{u}$ cannot be satisfied.

Let $\bar{c} = v^{-1}(u^{-1}(\underline{u}) + g(\bar{a}) + (\bar{\eta} - \underline{\eta})g'(\bar{a}))$. Then, all on the equilibrium consumptions are

no greater than \bar{c} . Hence, the terms in inequality (71) can be bounded as

$$\begin{aligned} \frac{g'(a(x))}{v'(c(x))} f(x) &\leq \frac{g'(\bar{a})}{v'(\bar{c})} f(x), \\ g''(a(x)) E \left[\frac{1}{v'(c(\eta))} 1_{\eta > x} \right] &\leq g''(\bar{a}) E \left[\frac{1}{v'(\bar{c})} 1_{\eta > x} \right] = g''(\bar{a}) \frac{\bar{F}(x)}{v'(\bar{c})}, \end{aligned}$$

which gives the claimed inequality.

E.3 Illustrations for Proposition 4

We now provide explicit conditions to verify the optimality of maximum effort in the three examples in Section 3.3.

Example 1. Let $u(x) = x$, $v(x) = x^\gamma$, $\gamma \in (0, 1]$. Consider the sub-case of $\underline{u} = 0$ and $g(a) = e^{Ga}$. As stated in the paper, the objective function is:

$$B(a) - e^{Ga/\gamma} E \left[(G\eta + 1)^{1/\gamma} \right].$$

Call a^{**} the solution of this problem. Proposition 4 proves that implementing a^{**} is optimal among all contracts (which need not implement a^{**}) if

$$\inf_{a \leq a^{**}} B'(a) f(\eta) \geq \lambda(a^{**}, \eta),$$

where $\lambda(a, \eta) = \max \left(0, \frac{\partial C[A]}{\partial a(\eta)} \right)$.

Inequality (72) establishes the bound

$$\frac{\partial C[A]}{\partial a(\eta)} \leq \Lambda(\bar{a}, \eta) = \frac{1}{\gamma} G e^{G\bar{a}/\gamma} (f(\eta) + G\bar{F}(\eta)) (1 + (\bar{\eta} - \underline{\eta})G)^{(1-\gamma)/\gamma}.$$

By Proposition 4, constant target effort a^{**} will be the optimum among all contracts (not necessarily requesting a constant effort) if

$$\forall a^{**}, \inf_{a \leq a^{**}} B'(a) \geq \frac{1}{\gamma} G e^{Ga^{**}/\gamma} \left(1 + G \sup_{\eta} \frac{\bar{F}(\eta)}{f(\eta)} \right) (1 + (\bar{\eta} - \underline{\eta})G)^{(1-\gamma)/\gamma}.$$

Example 2. Let $v(x) = \ln x$, $u(x) = e^{(1-\gamma)x}/(1-\gamma)$ for $\gamma > 0$, $\eta \sim N(0, \sigma^2)$ and $\underline{u} = u(\ln \underline{c})$. The contract specifying target effort a pays $c(\eta) = \underline{c} \exp(g'(a)\eta + g(a) - (1-\gamma)g'(a)^2\sigma^2/2)$.

The noise is unbounded here, so we will use equality (71) directly:

$$\begin{aligned}
\frac{\partial C[A]}{\partial a(x)} &= \underline{c} e^{(g(a)-(1-\gamma)g'(a)^2\sigma^2/2)} \left\{ g'(a) e^{g'(a)x} f(x) + \right. \\
&\quad \left. g''(a) E \left[e^{g'(a)\eta} 1_{\eta>x} \right] - g''(a) e^{(1-(1-\gamma)^2)g'(a)^2\sigma^2/2} E \left[e^{(1-\gamma)g'(a)\eta} 1_{\eta>x} \right] \right\} \\
&= \underline{c} e^{(g(a)-(1-\gamma)g'(a)^2\sigma^2/2)} \left\{ g'(a) e^{g'(a)x} \frac{1}{\sigma} \phi \left(\frac{x}{\sigma} \right) + \right. \\
&\quad \left. g''(a) e^{g'(a)^2\sigma^2/2} \left[\Phi \left(\frac{x}{\sigma} - (1-\gamma)\sigma g'(a) \right) - \Phi \left(\frac{x}{\sigma} - \sigma g'(a) \right) \right] \right\}.
\end{aligned}$$

Observing that

$$\Phi \left(\frac{x}{\sigma} - (1-\gamma)\sigma g'(a) \right) - \Phi \left(\frac{x}{\sigma} - \sigma g'(a) \right) = \gamma\sigma g'(a) \phi \left(\frac{x}{\sigma} \right) e^{-\xi^2/2 + \xi x/\sigma},$$

for some ξ between $(1-\gamma)\sigma g'(a)$ and $g'(a)\sigma$, we can obtain

$$\begin{aligned}
\frac{1}{f(x)} \frac{\partial C[A]}{\partial a(x)} &\leq \underline{c} e^{(g(a)-(1-\gamma)g'(a)^2\sigma^2/2)} g'(a) \left\{ e^{g'(a)x} + \gamma\sigma^2 g''(a) e^{g'(a)^2\sigma^2/2} e^{\xi x/\sigma} \right\} \\
&\leq \underline{c} e^{(g(a)-(1-\gamma)g'(a)^2\sigma^2/2)} g'(a) \left\{ e^{g'(a)x} + \right. \\
&\quad \left. \gamma\sigma^2 g''(a) e^{g'(a)^2\sigma^2/2} e^{g'(a)\max(x, (1-\gamma)x)} \right\}.
\end{aligned}$$

Let $\Lambda(a, x)$ denote the last upper bound. By Proposition 4, a^{**} will be the optimum if

$$\inf_{a \leq a^{**}} \partial_1 b(a, \eta, a^{**}) \geq \Lambda(a^{**}, \eta)$$

for all η .

Example 3. Let $v(x) = x$, and $u(x) = -e^{-\gamma x}$, for $\gamma > 0$, and $\eta \sim N(0, \sigma^2)$ as in HM. Similar to Example 2,

$$\frac{\partial C[A]}{\partial a(x)} = g'(a) \frac{1}{\sigma} \phi \left(\frac{x}{\sigma} \right) + g''(a) \left(\Phi \left(\frac{x}{\sigma} + \gamma\sigma g'(a) \right) - \Phi \left(\frac{x}{\sigma} \right) \right),$$

and

$$\frac{1}{f(x)} \frac{\partial C[A]}{\partial a(x)} \leq \Lambda(a, x) \equiv g'(a) + \gamma\sigma^2 g'(a) g''(a) e^{g'(a)\max(0, -\gamma x)}.$$

By Proposition 4, a^{**} will be the optimum if

$$\inf_{a \leq a^{**}} \partial_1 b(a, \eta, a^{**}) \geq \Lambda(a^{**}, \eta)$$

for all η .

F Quits and Firings

Our setup can be extended to accommodate quits and firing. We commence with the former. The agent now has an outside option available in each period t , and so the participation constraint in each period becomes $E_t[U_T] \geq \underline{u}_t$. As before, the principal wishes to implement $(a_t^*)_{t \leq T}$, and wishes to deter quitting. This can be achieved simply by increasing the constant K such that for all t , $E_t[U_T] \geq \underline{u}_t$. Under the conditions of Proposition 1, we can see that this is the only contract that ensures that. Economically, the agent receives rents because of his credible threat to leave in the interim periods. However, these rents only affect K , not the form of the contract. As in the core paper, if the benefit of effort is sufficiently high, maximum effort remains optimal.

We now turn to firings, considering $T = 2$ for simplicity and then discussing the generalizability to other T . Suppose that the principal wishes to fire the agent if $r_1 \in I_F$ and keep him if $r_1 \in I_F^c$, where I_F and I_F^c are disjoint intervals. Call r^F their common boundary, i.e. $\{r_1^F\} = \overline{I_F} \cap \overline{I_F^c}$. The next Proposition describes the contract.

Proposition 9 (*Contract with firing, $T = 2$*). *Under the conditions of Proposition 1 plus the option to fire, the following contract is optimal: (i) if $r_1 \in I_F$, the agent is fired, and receives a payoff $c = v^{-1}(g'(a_1^*)r_1 + K_1)$, (ii) if $r_1 \in I_F^c$, the agent remains employed, and receives a final payoff $c = v^{-1}(\sum_{t=1}^2 g'(a_t^*)r_t + K_2)$. The constants K_1 and K_2 are chosen such that the utility of the agent is continuous at $r_1 = r^F$, the cutoff return that triggers firing.*

Proof. (This is a sketch of the proof, as the arguments are similar to those in the main body of the paper). Define $\eta_1^F = r_1^F - a_1^*$, the cutoff noise that divides the regions of firing and not firing. For $\eta_1 \in \overset{\circ}{I}_{NF}^c$ (where $\overset{\circ}{I}$ is the interior of set I), by the logic of Proposition 1, very small deviations around a_1^* will still keep r_1 in $\overset{\circ}{I}_{NF}^c$ and so we require $c = v^{-1}(\sum_{t=1}^2 g'(a_t^*)r_t + K_{NF})$. For $\eta_1 \in \overset{\circ}{I}_F^c$, very small deviations around a_1^* will still keep r_1 in $\overset{\circ}{I}_F^c$, and so we require $c = v^{-1}(g'(a_1^*)r_1 + K_F)$ for some other constant. The utility should be continuous at r^F to preserve the IC. ■

Thus, the contract remains tractable even with the possibility of firing. This is because the intuition in the core model continues to hold – since the noise is observed before the action, the contract must provide sufficient incentives state-by-state and so the principal has little freedom in designing the contract. This contrasts with standard models in which the possibility of firing changes the contract significantly. The only degree of freedom for the principal is finding the domain I_F^c . As is standard, this will depend on the cost of finding another agent at $t = 2$. For instance, if the cost of finding a new employee are low, the domain of optimal firing might be large.

It is clear that the same logic would apply for $T > 2$. Suppose that the agent’s contract terminates at (a potentially return-dependent) time τ , with the same “tree” structure: at each time t , there is a monotone function $\Phi_t(r_1, \dots, r_t)$ such that the principal fires the agent if and

only if $\Phi_t(r_1, \dots, r_t) > 0$. Then, the compensation scheme has the following shape: if the agent works until τ , he receives:

$$c = v^{-1} \left(\sum_{t=1}^{\tau} g'(a_t^*) r_t + K_{\tau} \right) \quad (73)$$

for some constants K_1, \dots, K_T .

In addition, we can unify the two extensions of both quits and firings. Consider the firing model with $T = 2$. Suppose that the principal wishes to fire the agent if $r_1 \in I_F$, but also wishes to deter voluntary departures. Then, the contract is the one described in Proposition 9, but with K_1 and K_2 are simply set high enough such that the agent always receives at least his reservation utility.

G Proofs of Mathematical Lemmas

This section contains proofs of some of the mathematical lemmas featured in the appendices of the main paper.

Proof of Lemma 4 We thank Chris Evans for suggesting the proof strategy for this Lemma. We assume $a < b$.

We first prove the Lemma when $j(x) = 0 \forall x$. For a positive integer n , define $k_n = (b - a) / n$, and the function $r_n(x)$ as

$$r_n(x) = \begin{cases} \frac{f(x) - f(x - k_n)}{k_n} & \text{for } x \in [a + k_n, b] \\ 0 & \text{for } x \in [a, a + k_n). \end{cases}$$

We have for $x \in (a, b]$, $\liminf_{n \rightarrow \infty} r_n(x) \geq \liminf_{\varepsilon \downarrow 0} \frac{f(x) - f(x - \varepsilon)}{\varepsilon} \geq 0$.

Define $I_n = \int_a^b r_n(x) dx$. As $f+h$ is nondecreasing and k is C^1 , $\frac{f(x) - f(x - k_n)}{k_n} \geq \frac{-h(x) + h(x - k_n)}{k_n} \geq -\sup_{[a, b]} h'(x)$. Therefore, $r_n(x) \geq \min(0, -\sup_{[a, b]} h'(x)) \forall x$. Hence we can apply Fatou's lemma, which shows:

$$\liminf_{n \rightarrow \infty} I_n = \liminf_{n \rightarrow \infty} \int_a^b r_n(x) dx \geq \int_a^b \liminf_{n \rightarrow \infty} r_n(x) dx \geq 0.$$

Next, observe that $I_n = \int_{a+k_n}^b \frac{f(x) - f(x - k_n)}{k_n} dx$ consists of telescoping sums, so:

$$\begin{aligned} I_n &= \int_{b-k_n}^b \frac{f(x)}{k_n} dx - \int_a^{a+k_n} \frac{f(x)}{k_n} dx \\ &= f(b) - f(a) - \int_{b-k_n}^b \frac{f(b) - f(x)}{k_n} dx - \int_a^{a+k_n} \frac{f(x) - f(a)}{k_n} dx = f(b) - f(a) - B_n - A_n. \end{aligned}$$

We first minorize A_n . From condition (ii) of the Lemma, for any $\varepsilon > 0$, there is an $\eta > 0$, such that for $x \in [a, a + \eta]$, $f(x) - f(a) \geq -\varepsilon$. For n large enough such that $k_n \leq \eta$,

$$A_n = \int_a^{a+k_n} \frac{f(x) - f(a)}{k_n} dx \geq \int_a^{a+k_n} \frac{-\varepsilon}{k_n} dx = -\varepsilon,$$

and so $\liminf_{n \rightarrow \infty} A_n \geq 0$.

We next minorize B_n . Since $f'_-(b) \geq 0$ for every $\varepsilon > 0$, there exists a $\delta > 0$ s.t. for $x \in [b - \delta, b]$, $(f(b) - f(x)) / (b - x) \geq -\varepsilon$. Therefore, for n sufficiently large so that $k_n \leq \delta$,

$$B_n = \int_{b-k_n}^b \frac{f(b) - f(x)}{k_n} dx \geq \int_{b-k_n}^b \frac{(-\varepsilon)(b-x)}{k_n} dx = -\varepsilon \frac{k_n}{2},$$

and so $\liminf_{n \rightarrow \infty} B_n \geq 0$.

Finally, since $f(b) - f(a) = I_n + A_n + B_n$, we have

$$f(b) - f(a) = \liminf_{n \rightarrow \infty} (I_n + A_n + B_n) \geq \liminf_{n \rightarrow \infty} I_n + \liminf_{n \rightarrow \infty} A_n + \liminf_{n \rightarrow \infty} B_n \geq 0.$$

We now prove the general case. Define $F(x) = f(x) - \int_a^x j(t) dt$. Then, $F'_-(x) \geq 0$. By the above result, $F(b) - F(a) \geq 0$.

Proof of Lemma 5

Let $(y_n) \uparrow x$ be a sequence such that

$$f'_-(x) = \lim_{y_n \uparrow x} \frac{f(x) - f(y_n)}{x - y_n}.$$

We can further assume that $\lim_{n \rightarrow \infty} f(y_n)$ exists (if not, then we can choose a subsequence y_{n_k} such that $\lim_{n_k \rightarrow \infty} f(y_{n_k})$ exists and replace y_n by y_{n_k}).

If $\lim_{n \rightarrow \infty} f(y_n) = f(x)$, Then,

$$\begin{aligned} (h \circ f)'_-(x) &= \liminf_{y \uparrow x} \frac{h \circ f(x) - h \circ f(y)}{x - y} \\ &\leq \lim_{y_n \uparrow x} \frac{h \circ f(x) - h \circ f(y_n)}{x - y_n} \\ &= \lim_{y_n \uparrow x} \frac{h \circ f(x) - h \circ f(y_n)}{f(x) - f(y_n)} \frac{f(x) - f(y_n)}{x - y_n} \\ &= h'(f(x)) f'_-(x). \end{aligned}$$

If $\lim_{n \rightarrow \infty} f(y_n) < f(x)$, then $f'_-(x) = \infty$, since $h'(f(x)) > 0$, we still have $(h \circ f)'_-(x) \leq h'(f(x)) f'_-(x)$.

If $\lim_{n \rightarrow \infty} f(y_n) > f(x)$, then $(h \circ f)'_-(x) \leq \lim_{y_n \uparrow x} \frac{h \circ f(x) - h \circ f(y_n)}{x - y_n} = -\infty$, hence $(h \circ f)'_-(x) \leq h'(f(x)) f'_-(x)$.

On the other hand, suppose $(\hat{y}_n) \uparrow x$ be a sequence such that

$$(h \circ f)'_{-}(x) = \lim_{\hat{y}_n \uparrow x} \frac{h \circ f(x) - h \circ f(\hat{y}_n)}{x - \hat{y}_n},$$

and that $\lim_{n \rightarrow \infty} f(\hat{y}_n)$ exists. If $\lim_{n \rightarrow \infty} f(\hat{y}_n) = f(x)$, Then,

$$\begin{aligned} (h \circ f)'_{-}(x) &= \lim_{\hat{y}_n \uparrow x} \frac{h \circ f(x) - h \circ f(\hat{y}_n)}{x - \hat{y}_n} \\ &= \lim_{\hat{y}_n \uparrow x} \frac{h \circ f(x) - h \circ f(\hat{y}_n)}{f(x) - f(\hat{y}_n)} \frac{f(x) - f(\hat{y}_n)}{x - \hat{y}_n} \\ &= \lim_{\hat{y}_n \uparrow x} \frac{h \circ f(x) - h \circ f(\hat{y}_n)}{f(x) - f(\hat{y}_n)} \lim_{\hat{y}_n \uparrow x} \frac{f(x) - f(\hat{y}_n)}{x - \hat{y}_n} \\ &= h'(f(x)) \lim_{\hat{y}_n \uparrow x} \frac{f(x) - f(\hat{y}_n)}{x - \hat{y}_n} \mathfrak{5} \\ &\geq h'(f(x)) f'_{-}(x). \end{aligned}$$

Note that the existence of $\lim_{\hat{y}_n \uparrow x} \frac{h \circ f(x) - h \circ f(\hat{y}_n)}{x - \hat{y}_n}$ and $\lim_{\hat{y}_n \uparrow x} \frac{h \circ f(x) - h \circ f(\hat{y}_n)}{f(x) - f(\hat{y}_n)}$ guarantees the existence of $\lim_{\hat{y}_n \uparrow x} \frac{f(x) - f(\hat{y}_n)}{x - \hat{y}_n}$.

If $\lim_{n \rightarrow \infty} f(\hat{y}_n) < f(x)$, then $(h \circ f)'_{-}(x) = \infty \geq h'(f(x)) f'_{-}(x)$.

If $\lim_{n \rightarrow \infty} f(\hat{y}_n) > f(x)$, then $f'_{-}(x) \leq \lim_{\hat{y}_n \uparrow x} \frac{f(x) - f(\hat{y}_n)}{x - y_n} = -\infty \leq (h \circ f)'_{-}(x)$. Therefore, $(h \circ f)'_{-}(x) = h'(f(x)) f'_{-}(x)$.

Proof of Lemma 6

We use

$$\begin{aligned} (f + h)'_{-}(x) &= \liminf_{y \uparrow x} \frac{f(x) + h(x) - f(y) - h(y)}{x - y} = \liminf_{y \uparrow x} \left(\frac{f(x) - f(y)}{x - y} + \frac{h(x) - h(y)}{x - y} \right) \\ &\geq \liminf_{y \uparrow x} \frac{f(x) - f(y)}{x - y} + \liminf_{y \uparrow x} \frac{h(x) - h(y)}{x - y} = f'_{-}(x) + h'_{-}(x). \end{aligned}$$

When h is differentiable at x ,

$$(f + h)'_{-}(x) = \liminf_{y \uparrow x} \frac{f(x) - f(y)}{x - y} + \lim_{y \uparrow x} \frac{h(x) - h(y)}{x - y} = f'_{-}(x) + h'(x).$$

Proof of Lemma 7

We wish to prove that $E[h(X)] \geq E[h(Y)]$ for any concave function h . Define $I(\delta) = E[h(X + \delta(Y - X))]$ for $\delta \in [0, 1]$, so that

$$\begin{aligned} I''(\delta) &= E[h''(X + \delta(Y - X))(Y - X)^2] \leq 0 \\ I'(0) &= E[h'(X)(Y - X)] = E\left[h'(X) \left(\int_0^T \gamma_t dZ_t \right)\right], \end{aligned}$$

where $\gamma_t = \beta_t - \alpha_t$, and $\gamma_t \geq 0$ almost surely. We wish to prove $I(1) \leq I(0)$. Since I is concave, it is sufficient to prove that $I'(0) \leq 0$.

We next use some basic results from Malliavin calculus (see, e.g., Di Nunno, Oksendal and Proske (2008)). The integration by parts formula for Malliavin calculus yields:

$$I'(0) = \mathbb{E} \left[h'(X) \left(\int_0^T \gamma_t dZ_t \right) \right] = \mathbb{E} \left[\int_0^T (D_t h'(X)) \gamma_t dt \right],$$

where $D_t h'(X)$ is the Malliavin derivative of $h'(X)$ at time t . Since $(\alpha_s)_{s \in [0, T]}$ is deterministic. Therefore, the calculation of $D_t h'(X)$ is straightforward:

$$D_t h'(X) \equiv D_t h' \left(\int_0^T \alpha_s dZ_s \right) = h'' \left(\int_0^T \alpha_s dZ_s \right) \alpha_t = h''(X) \alpha_t.$$

Hence, we have:

$$I'(0) = \mathbb{E} \left[\int_0^T (D_t h'(X)) \gamma_t dt \right] = \mathbb{E} \left[\int_0^T h''(X) \alpha_t \gamma_t dt \right].$$

Since $h''(X) \leq 0$ (because h is concave), and α_t and γ_t are nonnegative, we have $h''(X) \alpha_t \gamma_t \leq 0$. Therefore, $I'(0) \leq 0$ as required.

References

- [1] DI NUNNO, G., B. OKSENDAL AND F. PROSKE (2008): *Malliavin Calculus for Lévy Processes with Applications to Finance*, Springer Verlag
- [2] HOLMSTROM, B. (1979): “Moral Hazard and Observability.” *Bell Journal of Economics* 10, 74-91
- [3] HOLMSTROM, B. (1982): “Moral Hazard in Teams.” *Bell Journal of Economics* 13, 324-340