

Online Appendix for “Tractability in Incentive Contracting” Not for Publication

Alex Edmans and Xavier Gabaix

April 22, 2011

A. Mathematical Preliminaries

This section derives some mathematical results that we use for the main proofs. This is a technical section that can be skipped.

A.1. Dispersion of Random Variables

We repeatedly use the “dispersive order” for random variables to show that IC constraints bind. Shaked and Shanthikumar (2007, Section 3.B) provide an excellent summary of known facts about this concept. This section provides a self-contained guide of the relevant results for our paper, as well as proving some new results. We extend results from Landsberger and Meilijson (1994), who use relative dispersion in another economic setting.

We commence by defining the notion of relative dispersion. Let X and Y denote two random variables with cumulative distribution functions F and G and corresponding right continuous inverses F^{-1} and G^{-1} . X is said to be less dispersed than Y if and only if $F^{-1}(\beta) - F^{-1}(\alpha) \leq G^{-1}(\beta) - G^{-1}(\alpha)$ whenever $0 < \alpha \leq \beta < 1$. This concept is location-free: X is less dispersed than Y if and only if it is less dispersed than $Y + z$, for any real constant z .

A basic property is the following result (Shaked and Shanthikumar (2007), p.151):

Lemma 2 *Let X be a random variable and f, h be functions such that $0 \leq f(y) - f(x) \leq h(y) - h(x)$ whenever $x \leq y$. Then $f(X)$ is less dispersed than $h(X)$.*

This result is intuitive: h magnifies differences to a greater extent than f , leading to more dispersion. We will also use the next two comparison lemmas.

Lemma 3 *Assume that X is less dispersed than Y and let f denote a weakly increasing function, h a weakly increasing concave function, and ϕ a weakly increasing convex function. Then:*

$$\begin{aligned} \mathbb{E}[f(X)] \geq \mathbb{E}[f(Y)] &\Rightarrow \mathbb{E}[h(f(X))] \geq \mathbb{E}[h(f(Y))] \\ \mathbb{E}[f(X)] \leq \mathbb{E}[f(Y)] &\Rightarrow \mathbb{E}[\phi(f(X))] \leq \mathbb{E}[\phi(f(Y))]. \end{aligned}$$

Proof. The first statement comes directly from Shaked and Shanthikumar (2007), Theorem 3.B.2, which itself is taken from Landsberger and Meilijson (1994). The second statement is derived from the first, applied to $\hat{X} = -X$, $\hat{Y} = -Y$, $\hat{f}(x) = -f(-x)$, $h(x) = -\phi(-x)$. It can

be verified directly (or via consulting Shaked and Shanthikumar (2007), Theorem 3.B.6) that \widehat{X} is less dispersed than \widehat{Y} . In addition, $E[\widehat{f}(\widehat{X})] \geq E[\widehat{f}(\widehat{Y})]$. Thus, $E[h(\widehat{f}(\widehat{X}))] \geq E[h(\widehat{f}(\widehat{Y}))]$. Substituting $h(\widehat{f}(\widehat{X})) = -\phi(f(X))$ yields $E[-\phi(f(X))] \geq E[-\phi(f(Y))]$. ■

Lemma 3 is intuitive: if $E[f(X)] \geq (\leq) E[f(Y)]$, applying a concave (convex) function should maintain the inequality. In addition, if $E[X] = E[Y]$, Lemma 3 implies that X second-order stochastically dominates Y . Hence, relative dispersion is a stronger concept than second-order stochastic dominance.

Lemma 3 allows us to prove Lemma 4 below, which states that the NIARA property of a utility function is preserved by adding a log-concave random variable to its argument.

Lemma 4 *Let u denote a utility function with NIARA and Y a log-concave random variable. Then, the utility function \widehat{u} defined by $\widehat{u}(x) \equiv E[u(x + Y)]$ exhibits NIARA.*

Proof. Consider two constants $a < b$ and a lottery Z independent from Y . Let C_a and C_b be the certainty equivalents of Z with respect to utility function \widehat{u} and evaluated at points a and b respectively, i.e. defined by

$$\widehat{u}(a + C_a) = E[u(a + Z)], \quad \widehat{u}(b + C_b) = E[u(b + Z)].$$

\widehat{u} exhibits NIARA if and only if $C_a \leq C_b$, i.e. the certainty equivalent increases with wealth. To prove that $C_a \leq C_b$, we make three observations. First, since u exhibits NIARA, there exists an increasing concave function h such that $u(a + x) = h(u(b + x))$ for all x . Second, because Y is log-concave, $Y + C_b$ is less dispersed than $Y + Z$ by Theorem 3.B.7 of Shaked and Shanthikumar (2007). Third, by definition of C_b and the independence of Y and Z , we have $E[u(b + Y + C_b)] = E[u(b + Y + Z)]$. Hence, we can apply Lemma 3, which yields $E[h(u(b + Y + C_b))] \geq E[h(u(b + Y + Z))]$, i.e.

$$E[u(a + Y + C_b)] \geq E[u(a + Y + Z)] = E[u(a + Y + C_a)] \text{ by definition of } C_a.$$

Thus we have $C_b \geq C_a$ as required. ■

A.2. Subderivatives

Since we cannot assume that the optimal contract is differentiable, we use the notion of subderivatives to allow for quasi first-order conditions in all cases. This concept is related to Krishna and Maenner's (2001) use of the subgradient, although the applications are quite different.

Definition 1 *For a point x and function f defined in a left neighborhood of x , we define the*

subderivative of f at x as:

$$\frac{d}{dx_-} f \equiv f'_-(x) \equiv \liminf_{y \uparrow x} \frac{f(x) - f(y)}{x - y}$$

This notion will prove useful since $f'_-(x)$ is well-defined for all functions f (with perhaps infinite values). We take limits “from below,” as we will often apply the subderivative at the highest feasible effort level \bar{a} . If f is left-differentiable at x , then $f'_-(x) = f'(x)$.

We use the following Lemma to allow us to integrate inequalities with subderivatives.

Lemma 5 *Assume that, over an interval I : (i) $f'_-(x) \geq j(x) \forall x$, for an continuous function $j(x)$ and (ii) there is a C^1 function h such that $f + h$ is nondecreasing. Then, for two points $a \leq b$ in I , $f(b) - f(a) \geq \int_a^b j(x) dx$.*

Proof. 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, $F(b) - F(a) \geq 0$. ■

Condition (ii) prevents $f(x)$ from exhibiting discontinuous downwards jumps, which would prevent integration.²⁰

The following Lemma is the chain rule for subderivatives.

Lemma 6 *Let x be a real number and f be a function defined in a left neighborhood of x . Suppose that function h is differentiable at $f(x)$, with $h'(f(x)) > 0$. Then, $(h \circ f)'_-(x) = h'(f(x)) f'_-(x)$.*

Proof. 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 $\lim_{k \rightarrow \infty} f(y_{n_k})$ 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 $(h \circ f)'_-(x) \leq h'(f(x)) f'_-(x)$.

²⁰For example, $f(x) = 1 \{x \leq 0\}$ satisfies condition (i) as $f'_-(x) = 0 \forall x$, but violates both condition (ii) and the conclusion of the Lemma, as $f(-1) > f(1)$.

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,

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

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, $(h \circ f)'_-(x) = h'(f(x)) f'_-(x)$. ■

In general, subderivatives typically follow the usual rules of calculus, with inequalities instead of equalities. One example is below.

Lemma 7 *Let x be a real number and f, h be functions defined in a left neighborhood of x . Then $(f + h)'_-(x) \geq f'_-(x) + h'_-(x)$. When h is differentiable at x , then $(f + h)'_-(x) = f'_-(x) + h'(x)$.*

Proof. 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 ,

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

■

B. Proofs

Throughout these proofs, we use tildes to denote random variables. For example, $\tilde{\eta}$ is the noise viewed as a random variable and η is a particular noise realization. $\mathbb{E}[f(\tilde{\eta})]$ denotes the expectation over all realizations of $\tilde{\eta}$ and $\mathbb{E}\left[\tilde{f}(\tilde{\eta})\right]$ denotes the expectation over all realizations of both x and a stochastic function \tilde{f} .

Proof of Theorem 1

Roadmap. We divide the proof in three parts. The first part shows that messages are redundant, so that we can restrict the analysis to contracts without messages. The second part proves the theorem considering only deterministic contracts and assuming that $a_t^* < \bar{a} \forall t$. This case requires weaker assumptions (see Remark 1). The third part, which is significantly more complex, rules out randomized contracts and allows for the target effort to be the maximum \bar{a} . Both these extensions require the concepts of subderivatives and dispersion from Appendix A.

1). Redundancy of Messages

In general, the contract can depend not only on the observed signals, but also messages M_t sent by the agent to the principal regarding the noise. The contract is now given by $\tilde{c}(r_1, \dots, r_T, M_1, \dots, M_T)$, and the agent's policy is now $(a, M) = (a_1, \dots, a_T, M_1, \dots, M_T)$. Let \mathbf{r} denote the vector (r_1, \dots, r_T) and define $\boldsymbol{\eta}$ and \mathbf{a} analogously. Define $\mathbf{g}(\mathbf{a}) = g(a_1) + \dots + g(a_T)$. Let $\tilde{V}_M(\mathbf{r}, \boldsymbol{\eta}) = v(\tilde{c}(\mathbf{r}, \boldsymbol{\eta}))$ denote the felicity given by a message-dependent contract if the agent reports $\boldsymbol{\eta}$ and the realized signals are \mathbf{r} . Under the revelation principle, we can restrict the analysis to mechanisms that induce the agent to truthfully report the noise $\boldsymbol{\eta}$. The incentive compatibility (IC) constraint is that the agent exerts effort \mathbf{a} and reports $\hat{\boldsymbol{\eta}} = \boldsymbol{\eta}$:

$$\forall \boldsymbol{\eta}, \forall \hat{\boldsymbol{\eta}}, \forall \mathbf{a}, \quad \mathbb{E}\left[u\left(\tilde{V}_M(\boldsymbol{\eta} + \mathbf{a}, \hat{\boldsymbol{\eta}}) - \mathbf{g}(\mathbf{a})\right)\right] \leq \mathbb{E}\left[u\left(\tilde{V}_M(\boldsymbol{\eta} + \mathbf{a}^*, \boldsymbol{\eta}) - \mathbf{g}(\mathbf{a}^*)\right)\right]. \quad (33)$$

The principal's problem is to minimize expected pay $\mathbb{E}\left[v^{-1}\left(\tilde{V}_M(\tilde{\boldsymbol{\eta}} + \mathbf{a}^*, \tilde{\boldsymbol{\eta}})\right)\right]$, subject to the IC constraint (33), and the agent's individual rationality (IR) constraint

$$\mathbb{E}\left[u\left(\tilde{V}_M(\tilde{\boldsymbol{\eta}} + \mathbf{a}^*, \tilde{\boldsymbol{\eta}}) - \mathbf{g}(\mathbf{a}^*)\right)\right] \geq \underline{u}. \quad (34)$$

Since $\mathbf{r} = \mathbf{r}^* \equiv \mathbf{a}^* + \boldsymbol{\eta}$ on the equilibrium path, the message-dependent contract is equivalent to $\tilde{V}_M(\mathbf{r}, \mathbf{r} - \mathbf{a}^*)$. We consider replacing this with a new contract $\tilde{V}(\mathbf{r})$, which only depends on the realized signal and not on any messages, and yields the same felicity as the corresponding message-dependent contract. Thus, the felicity it gives is defined by:

$$\tilde{V}(\mathbf{r}) = \tilde{V}_M(\mathbf{r}, \mathbf{r} - \mathbf{a}^*). \quad (35)$$

The IC and IR constraints for the new contract are given by:

$$\forall \eta, \forall a, \mathbb{E} \left[u \left(\tilde{V}(\mathbf{r}) - g(\mathbf{a}) \right) \right] \leq \mathbb{E} \left[u \left(\tilde{V}(\mathbf{r}^*) - g(\mathbf{a}^*) \right) \right], \quad (36)$$

$$\mathbb{E} \left[u \left(\tilde{V}(\mathbf{r}^*) - g(\mathbf{a}^*) \right) \right] \geq \underline{u}. \quad (37)$$

If the agent reports $\hat{\boldsymbol{\eta}} \neq \boldsymbol{\eta}$, he must take action \mathbf{a} such that $\boldsymbol{\eta} + \mathbf{a} = \hat{\boldsymbol{\eta}} + \mathbf{a}^*$. Substituting $\hat{\boldsymbol{\eta}} = \boldsymbol{\eta} + \mathbf{a} - \mathbf{a}^*$ into (33) and (34) indeed yields (36) and (37) above. Thus, the IC and IR constraints of the new contract are satisfied. Moreover, the new contract costs exactly the same as the old contract, since it yields the same felicity by (35). Hence, the new contract $\tilde{V}(\mathbf{r})$ induces incentive compatibility and participation at the same cost as the initial contract $\tilde{V}_M(\mathbf{r}, \boldsymbol{\eta})$ with messages, and so messages are not useful. The intuition is that \mathbf{a}^* is always exerted, so the principal can already infer $\boldsymbol{\eta}$ from the signal \mathbf{r} without requiring messages.

2). *Deterministic Contracts, in the case $a_t^* < \bar{a} \forall t$*

We will prove the Theorem by induction on T .

2a). *Case $T = 1$.* Dropping the time subscript for brevity, the incentive compatibility (IC) constraint is:

$$\forall \eta, \forall a : V(a + \eta) - g(a) \leq V(a^* + \eta) - g(a^*) \quad (38)$$

Fix $a^* \in (\underline{a}, \bar{a})$ and $r \in (a^* + \underline{\eta}, a^* + \bar{\eta})$, and consider some r' . We pick a, a', η, η' such that:

$$\begin{aligned} r &= a^* + \eta & r' &= a + \eta \\ r &= a' + \eta' & r' &= a^* + \eta' \end{aligned}$$

i.e., pick

$$\begin{aligned} \eta &= r - a^* & a &= a^* + r' - r \\ \eta' &= r' - a^* & a' &= a^* + r - r' \end{aligned}$$

This implies $\eta \in (\underline{\eta}, \bar{\eta})$. Also, by continuity of the function $r' \mapsto (a, a', \eta')$, there is a $\varepsilon > 0$ such that $\forall r' \in (r - \varepsilon, r + \varepsilon)$, $a, a' \in (\underline{a}, \bar{a})$ and $\eta' \in (\underline{\eta}, \bar{\eta})$. From now on, consider only $r' \in (r - \varepsilon, r + \varepsilon)$. We can apply equation (38) to (a, η) and then to (a', η') . This gives

$$V(r') - g(a^* + r' - r) \leq V(r) - g(a^*)$$

and then

$$V(r) - g(a^* + r - r') \leq V(r') - g(a^*).$$

Hence,

$$g(a^*) - g(a^* + r' - r) \leq V(r) - V(r') \leq g(a^* + r - r') - g(a^*). \quad (39)$$

We first consider $r > r'$. Dividing through by $r - r'$ yields:

$$\frac{g(a^*) - g(a^* + r' - r)}{r - r'} \leq \frac{V(r) - V(r')}{r - r'} \leq \frac{g(a^* + r - r') - g(a^*)}{r - r'}. \quad (40)$$

Taking the limit $r' \uparrow r$, the first and third terms of (40) converge to $g'(a^*)$. Therefore, the left derivative $V'_{left}(r)$ exists, and equals $g'(a^*)$. Second, consider $r < r'$. Dividing (39) through by $r - r'$, and taking the limit $r' \downarrow r$ shows that the right derivative $V'_{right}(r)$ exists, and equals $g'(a^*)$. Therefore,

$$V'(r) = g'(a^*). \quad (41)$$

Since r has interval support²¹, we can integrate to obtain, for some integration constant K :

$$V(r) = g'(a^*)r + K. \quad (42)$$

2b). *If the Theorem holds for T , it holds for $T + 1$.* This part is as in the main text.

Note that the above proof (for deterministic contracts where $a_t^* < \bar{a}$) does not require log-concavity of η_t , nor that u satisfies NIARA. This is because formula (6) is satisfied by all an incentive contracts for signals $(r_t)_{t=1\dots T}$ in $X_{t=1}^T \left(\underline{\eta}_t + a_t^*, \bar{\eta}_t + a_t^* \right)$, i.e. that are almost surely observed if the agent follows the recommended policy. These assumptions are only required for the general proof, where other contracts (e.g. randomized ones) are also incentive compatible, to show that they are costlier than contract (6).

3). *General Proof*

We no longer restrict a_t^* to be in the interior of \mathcal{A} , and allow for randomized contracts. We wish to prove the following statement Σ_T by induction on integer T :

Statement Σ_T . *Consider a utility function u with NIARA, independent random variables $\tilde{r}_1, \dots, \tilde{r}_T$ where $\tilde{r}_2, \dots, \tilde{r}_T$ are log-concave, and a sequence of nonnegative numbers $g'(a_1^*), \dots, g'(a_T^*)$. Consider the set of (potentially randomized) contracts $\tilde{V}(r_1, \dots, r_T)$ such that (i) $\mathbb{E} \left[u \left(\tilde{V}(\tilde{r}_1, \dots, \tilde{r}_T) \right) \right] \geq \underline{u}$; (ii) $\forall t = 1\dots T$,*

$$\frac{d}{d\varepsilon_-} \mathbb{E} \left[u \left(\tilde{V}(\tilde{r}_1, \dots, \tilde{r}_t + \varepsilon, \dots, \tilde{r}_T) \right) \mid \tilde{r}_1, \dots, \tilde{r}_t \right]_{\varepsilon=0} \geq g'(a_t^*) \mathbb{E} \left[u' \left(\tilde{V}(\tilde{r}_1, \dots, \tilde{r}_t, \dots, \tilde{r}_T) \right) \mid \tilde{r}_1, \dots, \tilde{r}_t \right] \quad (43)$$

and (iii) $\forall t = 1\dots, T$, $\mathbb{E} \left[u \left(\tilde{V}(\tilde{r}_1, \dots, \tilde{r}_t, \dots, \tilde{r}_T) \right) \mid \tilde{r}_1, \dots, \tilde{r}_t \right]$ is nondecreasing in \tilde{r}_t .

In this set, for any increasing and convex cost function ϕ , $\mathbb{E}[\phi(V(\tilde{r}_1, \dots, \tilde{r}_T))]$ is minimized with contract: $V^0(r_1, \dots, r_T) = \sum_{t=1}^T g'(a_t^)r_t + K$, where K is a constant that makes the participation constraint (i) bind.*

Condition (ii) is the local IC constraint, for deviations from below.

²¹The model could be extended to allowing non-interval support: if the domain of r was a union of disjoint intervals, we would have a different integration constant K for each interval.

We first consider the case of deterministic contracts, and then show that randomized contracts are costlier. We use the notation $E_t[\cdot] = E[\cdot \mid \tilde{r}_1, \dots, \tilde{r}_t]$ to denote the expectation based on time- t information.

3a). *Deterministic Contracts*

The key difference from the proof in 1) is that we now must allow for $a_t^* = \bar{a}$.

3ai). *Proof of Statement Σ_T when $T = 1$.*

(43) becomes $\frac{d}{d\varepsilon_-} u(V(r + \varepsilon))|_{\varepsilon=0} \geq g'(a_1^*) u'(V(r))$. Applying Lemma 6 to $h = u^{-1}$ yields:

$$V'_-(r) \geq g'(a^*). \quad (44)$$

It is intuitive that (44) should bind, as this minimizes the variability in the agent's pay and thus constitutes efficient risk-sharing. We now prove that this is indeed the case; to simplify exposition, we normalize $g(a^*) = 0$ w.l.o.g.²² If constraint (44) binds, the contract is $V^0(r) = g'(a^*)r + K$, where K satisfies $E[u(g'(a^*)r + K)] = \underline{u}$. We wish to show that any other contract $V(r)$ that satisfies (44) is weaklier costlier.

By assumption (iii) in Statement Σ_1 , V is nondecreasing. We can therefore apply Lemma 5 to equation (44), where condition (ii) of the Lemma is satisfied by $h(r) \equiv 0$. This implies that for $r \leq r'$, $V(r') - V(r) \geq g'(a^*)(r' - r) = V^0(r') - V^0(r)$. Thus, using Lemma 2, $V(\tilde{r})$ is more dispersed than $V^0(\tilde{r})$.

Since V must also satisfy the participation constraint, we have:

$$E[u(V(\tilde{r}))] \geq \underline{u} = E[u(V^0(\tilde{r}))]. \quad (45)$$

Applying Lemma 3 to the convex function $\phi \circ u^{-1}$ and inequality (45), we have:

$$E[\phi \circ u^{-1} \circ u(V(\tilde{r}))] \geq E[\phi \circ u^{-1} \circ u(V^0(\tilde{r}))],$$

i.e. $E[\phi(V(\tilde{r}))] \geq E[\phi(V^0(\tilde{r}))]$. The expected cost of V^0 is weakly less than for V . Hence, the contract V^0 is cost-minimizing.

We note that this last part of the reasoning underpins item 2 in Section 1.3, the extension to a risk-averse principal. Suppose that the principal wants to minimize $E[w(c)]$, where w is an increasing and concave function, rather than $E[c]$. Then, the above contract is optimal if $w \circ v^{-1} \circ u^{-1}$ is convex, i.e. $u \circ v \circ w^{-1}$ is concave. This requires w to be “not too concave,” i.e. the agent to be not too risk-averse.

Finally, we verify that the contract V^0 satisfies the global IC constraint. The agent's objective function becomes $u(g'(a^*)(a + \eta) - g(a))$. Since $g(a)$ is convex, the argument of $u(\cdot)$ is concave. Hence, the first-order condition gives the global optimum.

²²Formally, this can be achieved by replacing the utility function $u(x)$ by $u^{new}(x) = u(x - g(a^*))$ and the cost function $g(a)$ by $g^{new}(a) = g(a) - g(a^*)$, so that $u(x - g(a)) = u^{new}(x - g^{new}(a))$.

3aii). *Proof that if Statement Σ_T holds for T , it holds for $T + 1$.* We define a new utility function \hat{u} as follows:

$$\hat{u}(x) = \mathbb{E} [u(x + g'(a_{T+1}^*) \tilde{r}_{T+1})]. \quad (46)$$

Since \tilde{r}_{T+1} is log-concave, $g'(a_{T+1}^*) \tilde{r}_{T+1}$ is also log-concave. From Lemma 4, \hat{u} has the same NIARA property as u .

For each $\tilde{r}_1, \dots, \tilde{r}_T$, we define $k(\tilde{r}_1, \dots, \tilde{r}_T)$ as the solution to equation (47) below:

$$\hat{u}(k(\tilde{r}_1, \dots, \tilde{r}_T)) = \mathbb{E}_T [u(V(\tilde{r}_1, \dots, \tilde{r}_{T+1}))]. \quad (47)$$

k represents the expected felicity from contract V based on all signals up to time T .

The goal is to show that any other contract $V \neq V^0$ is weakly costlier. To do so, we wish to apply Statement Σ_T for utility function \hat{u} and contract k . The first step is to show that, if Conditions (i)-(iii) hold for utility function u and contract V at time $T + 1$, they also hold for \hat{u} and k at time T , thus allowing us to apply the Statement for these functions.

Taking expectations of (47) over $\tilde{r}_1, \dots, \tilde{r}_T$ yields:

$$\mathbb{E} [\hat{u}(k(\tilde{r}_1, \dots, \tilde{r}_T))] = \mathbb{E} [u(V(\tilde{r}_1, \dots, \tilde{r}_{T+1}))] \geq \underline{u}, \quad (48)$$

where the inequality comes from Condition (i) for utility function u and contract V at time $T + 1$. Hence, Condition (i) holds for utility function \hat{u} and contract k at time t . In addition, it is immediate that $\mathbb{E} [\hat{u}(k(\tilde{r}_1, \dots, \tilde{r}_T)) \mid \tilde{r}_1, \dots, \tilde{r}_t]$ is nondecreasing in \tilde{r}_t . (Condition (iii)). We thus need to show that Condition (ii) is satisfied. Since equation (43) holds for $t = T + 1$, we have

$$\frac{d}{d\varepsilon_-} u(V(\tilde{r}_1, \dots, \tilde{r}_T, \tilde{r}_{T+1} + \varepsilon)) \geq g'(a_{T+1}^*) u'[V(\tilde{r}_1, \dots, \tilde{r}_{T+1})].$$

Applying Lemma 6 with function u yields:

$$\frac{dV}{dr_{T+1-}}(r_1, \dots, r_{T+1}) \geq g'(a_{T+1}^*). \quad (49)$$

Hence, using Lemmas 2 and 5, we see that conditional on $\tilde{r}_1, \dots, \tilde{r}_T$, $V(\tilde{r}_1, \dots, \tilde{r}_{T+1})$ is more dispersed than $k(\tilde{r}_1, \dots, \tilde{r}_T) + g'(a_{T+1}^*) \tilde{r}_{T+1}$. Using (46), we can rewrite (47) as

$$\mathbb{E}_T [u(k(\tilde{r}_1, \dots, \tilde{r}_T) + g'(a_{T+1}^*) \tilde{r}_{T+1})] = \mathbb{E}_T [u(V(\tilde{r}_1, \dots, \tilde{r}_{T+1}))].$$

Since u exhibits NIARA, $-u''(x)/u'(x)$ is nonincreasing in x . This is equivalent to $u' \circ u^{-1}$ being weakly convex. We can thus apply Lemma 3 to yield:

$$\begin{aligned} \mathbb{E}_T [u' \circ u^{-1} \circ u(V(\tilde{r}_1, \dots, \tilde{r}_{T+1}))] &\geq \mathbb{E}_T [u' \circ u^{-1} \circ u(k(\tilde{r}_1, \dots, \tilde{r}_T) + g'(a_{T+1}^*) \tilde{r}_{T+1})], \text{ i.e.} \\ \mathbb{E}_T [u'(V(\tilde{r}_1, \dots, \tilde{r}_{T+1}))] &\geq \mathbb{E}_T [\hat{u}'(k(\tilde{r}_1, \dots, \tilde{r}_T))]. \end{aligned} \quad (50)$$

Applying definition (47) to the LHS of Condition (ii) for $T + 1$ yields, with $t = 1 \dots T$,

$$\frac{d}{d\varepsilon_-} \mathbf{E}_t [\widehat{u}(k(\tilde{r}_1, \dots, \tilde{r}_t + \varepsilon, \dots, \tilde{r}_T))]_{|\varepsilon=0} \geq g'(a_t^*) \mathbf{E} [u'(V(\tilde{r}_1, \dots, \tilde{r}_t, \dots, \tilde{r}_{T+1})) | \tilde{r}_1, \dots, \tilde{r}_t]$$

Taking expectations of equation (50) at time t and substituting into the RHS of the above equation yields:

$$\begin{aligned} \frac{d}{d\varepsilon_-} \mathbf{E}_t [\widehat{u}(k(\tilde{r}_1, \dots, \tilde{r}_t + \varepsilon, \dots, \tilde{r}_T))] &= \frac{d}{d\varepsilon_-} \mathbf{E}_t [u(V(\tilde{r}_1, \dots, \tilde{r}_t + \varepsilon, \dots, \tilde{r}_{T+1}))]_{|\varepsilon=0} \\ &\geq g'(a_t^*) \mathbf{E}_t [\widehat{u}'(k(\tilde{r}_1, \dots, \tilde{r}_T))] . \end{aligned}$$

Hence the IC constraint holds for contract $k(\tilde{r}_1, \dots, \tilde{r}_T)$ and utility function \widehat{u} at time T , and so Condition (ii) of Statement Σ_T is satisfied. We can therefore apply Statement Σ_T at T to contract $k(r_1, \dots, r_T)$, utility function \widehat{u} and cost function $\widehat{\phi}$ defined by:

$$\widehat{\phi}(x) \equiv \mathbf{E} [\phi(x + g'(a_{T+1})\tilde{r}_{T+1})] . \quad (51)$$

We observe that the contract $V^0 = \sum_{t=1}^{T+1} g'(a_t^*) r_t + K$ satisfies:

$$\mathbf{E} \left[\widehat{u} \left(\sum_{t=1}^T g'(a_t^*) r_t + K \right) \right] = \mathbf{E} \left[u \left(\sum_{t=1}^{T+1} g'(a_t^*) r_t + K \right) \right] = \underline{u} .$$

Therefore, applying Statement Σ_T to k , \widehat{u} and $\widehat{\phi}$ implies:

$$C_k = \mathbf{E} [\widehat{\phi}(k(\tilde{r}_1, \dots, \tilde{r}_T))] \geq C_{V^0} = \mathbf{E} \left[\phi \left(\sum_{t=1}^{T+1} g'(a_t^*) \tilde{r}_t + K \right) \right] . \quad (52)$$

Using equation (51) yields:

$$C_k = \mathbf{E} [\phi(k(\tilde{r}_1, \dots, \tilde{r}_T) + g'(a_{T+1})\tilde{r}_{T+1})] \geq C_{V^0} = \mathbf{E} \left[\phi \left(\sum_{t=1}^{T+1} g'(a_t^*) \tilde{r}_t + K \right) \right] .$$

Finally, we compare the cost of contract $k(r_1, \dots, r_T) + g'(a_{T+1})\tilde{r}_{T+1}$ to the cost of the original contract $V(r_1, \dots, r_{T+1})$. Since equation (47) is satisfied, we can apply Lemma 3 to the convex function $\phi \circ u^{-1}$ and the random variable \tilde{r}_{T+1} to yield

$$\begin{aligned} \mathbf{E}_t [\phi(V(\tilde{r}_1, \dots, \tilde{r}_{T+1}))] &\geq \mathbf{E}_t [\phi(k(\tilde{r}_1, \dots, \tilde{r}_T) + g'(a_{T+1}^*)\tilde{r}_{T+1})] \\ \mathbf{E} [\phi(V(\tilde{r}_1, \dots, \tilde{r}_{T+1}))] &\geq \mathbf{E} [\phi(k(\tilde{r}_1, \dots, \tilde{r}_T) + g'(a_{T+1}^*)\tilde{r}_{T+1})] = C_k \geq C_{V^0} . \end{aligned}$$

where the final inequality comes from (52). Hence the cost of contract k is weakly greater than the cost of contract V^0 . This concludes the proof for $T + 1$.

3b). *Optimality of Deterministic Contracts*

Consider a randomized contract $\tilde{V}(r_1, \dots, r_T)$ and define the “certainty equivalent” contract \bar{V} by:

$$u(\bar{V}(r_1, \dots, r_T)) \equiv E_T \left[u \left(\tilde{V}(r_1, \dots, r_T) \right) \right]. \quad (53)$$

We wish to apply Statement Σ_T (which we have already proven for deterministic contracts) to contract \bar{V} , and so must verify that its three conditions are satisfied.

From the above definition, we obtain

$$E \left[u(\bar{V}(\tilde{r}_1, \dots, \tilde{r}_T)) \right] = E \left[u \left(\tilde{V}(\tilde{r}_1, \dots, \tilde{r}_T) \right) \right] \geq \underline{u},$$

i.e., \bar{V} satisfies the participation constraint (34). Hence, Condition (i) holds. Also, it is clear that Condition (iii) holds for \bar{V} , given it holds for \tilde{V} . We thus need to show that Condition (ii) is also satisfied. Applying Jensen’s inequality to equation (53) and the function $u' \circ u^{-1}$ (which is convex since u exhibits NIARA) yields: $u'(\bar{V}(r_1, \dots, r_T)) \leq E_T \left[u' \left(\tilde{V}(r_1, \dots, r_T) \right) \right]$. We apply this to $r_t = \tilde{r}_t$ for $t = 1 \dots T$ and take expectations to obtain

$$E_t \left[u' \left(\tilde{V}(\tilde{r}_1, \dots, \tilde{r}_T) \right) \right] \geq E_t \left[u'(\bar{V}(\tilde{r}_1, \dots, \tilde{r}_T)) \right]. \quad (54)$$

Applying definition (53) to the left-hand side of (43) yields:

$$\frac{d}{d\varepsilon_-} E_t \left[u(\bar{V}(\tilde{r}_1, \dots, \tilde{r}_t + \varepsilon, \dots, \tilde{r}_T)) \right]_{|\varepsilon=0} \geq g'(a_t^*) E_t \left[u' \left(\tilde{V}(\tilde{r}_1, \dots, \tilde{r}_t, \dots, \tilde{r}_T) \right) \right].$$

and using (54) yields:

$$\frac{d}{d\varepsilon_-} E_t \left[u(\bar{V}(\tilde{r}_1, \dots, \tilde{r}_t + \varepsilon, \dots, \tilde{r}_T)) \right]_{|\varepsilon=0} \geq g'(a_t^*) E_t \left[u'(\bar{V}(\tilde{r}_1, \dots, \tilde{r}_t, \dots, \tilde{r}_T)) \right].$$

Condition (ii) of Statement Σ_T therefore holds for \bar{V} . We can therefore apply Statement Σ_T to show that V^0 has a weakly lower cost than \bar{V} . We next show that the cost of \bar{V} is weakly less than the cost of \tilde{V} . Applying Jensen’s inequality to (53) and the convex function $\phi \circ u^{-1}$ yields: $\phi(\bar{V}(r_1, \dots, r_T)) \leq E \left[\phi \left(\tilde{V}(r_1, \dots, r_T) \right) \right]$. We apply this to $r_t = \tilde{r}_t$ for $t = 1 \dots T$ and take expectations over the distribution of \tilde{r}_t to obtain:

$$\phi(\bar{V}(\tilde{r}_1, \dots, \tilde{r}_T)) \leq E \left[\phi \left(\tilde{V}(\tilde{r}_1, \dots, \tilde{r}_T) \right) \right].$$

Hence \bar{V} has a weakly lower cost than \tilde{V} . Therefore, V^0 has a weakly lower cost than \tilde{V} . This proves the Statement for randomized contracts.

3c). *Main Proof.* Having proven Statement Σ_T , we now turn to the main proof of Theorem

1. The value of the signal on the equilibrium path is given by $\tilde{r}_t \equiv a_t^* + \tilde{\eta}_t$. We define

$$\bar{u}(x) \equiv u \left(x - \sum_{s=1}^T g(a_s^*) \right). \quad (55)$$

We seek to use Statement Σ_T applied to function \bar{u} and random variable \tilde{r}_t , and thus must verify that its three conditions are satisfied. Since $\mathbb{E} \left[\bar{u} \left(\tilde{V}(\tilde{r}_1, \dots, \tilde{r}_T) \right) \right] \geq \underline{u}$, Condition (i) holds.

The IC constraint for time t is:

$$0 \in \arg \max_{\varepsilon} \mathbb{E}_t u \left(\tilde{V}(a_1^* + \tilde{\eta}_1, \dots, a_t^* + \tilde{\eta}_t + \varepsilon, \dots, a_T^* + \tilde{\eta}_T) - g(a_t^* + \varepsilon) - \sum_{s=1 \dots T, s \neq t} g(a_s^*) \right),$$

i.e.

$$0 \in \arg \max_{\varepsilon} \mathbb{E}_t u \left(\tilde{V}(\tilde{r}_1, \dots, \tilde{r}_t + \varepsilon, \dots, \tilde{r}_T) - g(a_t^* + \varepsilon) - \sum_{s=1 \dots T, s \neq t} g(a_s^*) \right). \quad (56)$$

We note that, for a function $f(\varepsilon)$, $0 \in \arg \max_{\varepsilon} f(\varepsilon)$ implies that for all $\varepsilon < 0$, $(f(0) - f(\varepsilon)) / (-\varepsilon) \geq 0$, hence, taking the $\liminf_{\varepsilon \uparrow 0}$, we obtain $\frac{d}{d\varepsilon_-} f'(\varepsilon)|_{\varepsilon=0} \geq 0$. Call $X(\varepsilon)$ the argument of u in equation (56). Applying this result to (56), we find: $\frac{d}{d\varepsilon_-} \mathbb{E}_t u(X(\varepsilon))|_{\varepsilon=0} \geq 0$.

Using Lemma 6, we find $\mathbb{E}_t \left[u'(X(0)) \left(\frac{d}{d\varepsilon_-} X(\varepsilon)|_{\varepsilon=0} \right) \right] \geq 0$. Using Lemma 7, $\frac{d}{d\varepsilon_-} X(\varepsilon)|_{\varepsilon=0} = \frac{d}{d\varepsilon_-} \tilde{V}(\tilde{r}_1, \dots, \tilde{r}_t + \varepsilon, \dots, \tilde{r}_T) - g'(a_t^*)$, hence we obtain:

$$\mathbb{E}_t \left[u'(X(0)) \left(\frac{d}{d\varepsilon_-} \tilde{V}(\tilde{r}_1, \dots, \tilde{r}_t + \varepsilon, \dots, \tilde{r}_T) - g'(a_t^*) \right) \right] \geq 0.$$

Using again Lemma 6, this can be rewritten:

$$\frac{d}{d\varepsilon_-} \mathbb{E}_t \left[u \left(\tilde{V}(\tilde{r}_1, \dots, \tilde{r}_t + \varepsilon, \dots, \tilde{r}_T) - \sum_{s=1 \dots T} g(a_s^*) \right) \right]_{\varepsilon=0} \geq g'(a_t^*) \mathbb{E}_t [u'(X(0))],$$

i.e., using the notation (55),

$$\frac{d}{d\varepsilon_-} \mathbb{E}_t \left[\bar{u} \left(\tilde{V}(\tilde{r}_1, \dots, \tilde{r}_t + \varepsilon, \dots, \tilde{r}_T) \right) \right]_{\varepsilon=0} \geq g'(a_t^*) \mathbb{E}_t \left[\bar{u}' \left(\tilde{V}(\tilde{r}_1, \dots, \tilde{r}_t, \dots, \tilde{r}_T) \right) \right].$$

Therefore, Condition (ii) of Statement Σ_T holds.

Finally, we verify Condition (iii). Apply (56) to signal r_t and deviation $\varepsilon < 0$. We obtain:

$$\begin{aligned} & \mathbb{E}_t \left[u \left(\tilde{V}(\tilde{r}_1, \dots, \tilde{r}_t + \varepsilon, \dots, \tilde{r}_T) - \sum_{s=1 \dots T} g(a_s^*) \right) \right] \\ & \geq \mathbb{E}_t \left[u \left(\tilde{V}(\tilde{r}_1, \dots, \tilde{r}_t + \varepsilon, \dots, \tilde{r}_T) - g(a_t^* + \varepsilon) - \sum_{s=1 \dots T, s \neq t} g(a_s^*) \right) \right] \\ & \geq \mathbb{E}_t \left[u \left(\tilde{V}(r_1, \dots, r_t + \varepsilon, \dots, r_T) - g(a_t^*) - \sum_{s=1 \dots T, s \neq t} g(a_s^*) \right) \right], \end{aligned}$$

so Condition (iii) holds for contract \tilde{V} and utility function \bar{u} .

We can now apply Statement Σ_T to contract \tilde{V} and function \bar{u} , to prove that any globally IC contract is weakly costlier than contract $V^0 = \sum_{t=1}^T g'(a_t^*) r_t + K$. Moreover, it is clear that V^0 satisfies the global IC conditions in equation (56). Thus, V^0 is the cheapest contract that satisfies the global IC constraint.

Proof of Remark 1

Conditionally on $(\eta_t)_{t \leq T+1}$, we must have:

$$a_{T+1}^* \in \arg \max_{a_{T+1}} u \left(V(a_1^* + \eta_1, \dots, a_{T+1}^* + \eta_{T+1}) - g(a_{T+1}) - \sum_{t \neq T+1} g(a_t^*) \right).$$

Using the proof of Theorem 1 with $T = 1$, this implies that, for r_{T+1} in the interior of the support of \tilde{r}_{T+1} (given $(r_t)_{t \leq T}$), $V(r_1, \dots, r_{T+1})$ can be written:

$$V(r_1, \dots, r_{T+1}) = K_T(r_1, \dots, r_T) + g'(a_{T+1}^*) r_{T+1},$$

for some function $K_T(r_1, \dots, r_T)$. Next, consider the problem of implementing action a_T^* at time T . We require that, for all $(\eta_t)_{t \leq T}$,

$$a_T^* \in \arg \max_{a_T} \mathbb{E}_T \left[u \left(K_T(a_1^* + \eta_1, \dots, a_T^* + \eta_T) + g'(a_{T+1}^*) (\eta_{T+1} + a_{T+1}^*) - g(a_T) - \sum_{t \neq T} g(a_t^*) \right) \right].$$

This can be rewritten

$$a_T^* \in \arg \max_{a_T} \hat{u}(K_T(a_1^* + \eta_1, \dots, a_T^* + \eta_T) - g(a_T)),$$

where $\hat{u}(x) \equiv \mathbb{E} \left[u \left(x + g'(a_{T+1}^*) (\eta_{T+1} + a_{T+1}^*) - \sum_{t \neq T} g(a_t^*) \right) \mid \eta_1, \dots, \eta_T \right]$.

Using the same arguments as above for $T + 1$, that implies that, for r_T in the interior of the

support of \tilde{r}_T (given $(r_t)_{t \leq T-1}$) we can write:

$$K_T(r_1, \dots, r_T) = K_{T-1}(r_1, \dots, r_{T-1}) + g'(a_T^*) r_T$$

for some function $K_{T-1}(r_1, \dots, r_{T-1})$. Proceeding by induction, we see that this implies that we can write, for $(r_t)_{t \leq T+1}$ in the interior of the support of $(\tilde{r}_t)_{t \leq T+1}$,

$$V_{T+1}(r_1, \dots, r_{T+1}) = \sum_{t=1}^{T+1} g'(a_t^*) r_t + K_0,$$

for some constant K_0 . This yields the “necessary” first part of the Proposition.

The converse part of the Proposition is immediate. Given the proposed contract, the agent faces the decision:

$$\max_{(a_t)_{t \leq T}} \mathbb{E} \left[u \left(\sum_{t=1}^T g'(a_t^*) a_t - g(a_t) + \sum_{t=1}^T g'(a_t^*) \eta_t \right) \right],$$

which is maximized pointwise when $g'(a_t^*) a_t - g(a_t)$ is maximized. This in turn requires $a_t = a_t^*$.

Proof of Proposition 1

We shall use the following purely mathematical Lemma.

Lemma 8 *Consider a standard Brownian process Z_t with filtration \mathcal{F}_t , a deterministic non-negative process α_t , an \mathcal{F}_t -adapted process β_t , $T \geq 0$, $X = \int_0^T \alpha_t dZ_t$, and $Y = \int_0^T \beta_t dZ_t$. Suppose that almost surely, $\forall t \in [0, T]$, $\alpha_t \leq \beta_t$. Then X second-order stochastically dominates Y .*

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

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

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. ■

Lemma 8 is intuitive: since $\beta_t \geq \alpha_t \geq 0$, it makes sense that Y is more volatile than X .

To derive the IC constraint, we use the methodology introduced by Sannikov (2008). We observe that the term $\int_0^T \mu_t dt$ induces a constant shift, so w.l.o.g. we can assume $\mu_t = 0 \forall t$.

For an arbitrary adapted policy function $a = (a_t)_{t \in [0, T]}$, let Q^a denote the probability measures induced by a . Then, $Z_t^a = \int_0^t (dr_s - a_s ds) / \sigma_s$ is a Brownian motion under Q^a , and $Z_t^{a^*} = \int_0^t (dr_s - a_s^* ds) / \sigma_s$ is a Brownian under Q^{a^*} , where a^* is the policy $(a_t^*)_{t \in [0, T]}$.

Recall that, if the agent exerts policy a^* , then $r_t = \int_0^t a_s^* ds + \sigma_s dZ_s$. We define $v_T = v(c)$. By the martingale representation theorem (Karatzas and Shreve (1991), p. 182) applied to process $v_t = E_t[v_T]$ for $t \in [0, T]$, we can write: $v_T = \int_0^T \theta_t (dr_t - a_t^* dt) + v_0$ for some constant v_0 and a process θ_t adapted to the filtration induced by $(r_s)_{s \leq t}$.

We proceed in two steps.

1) We show that policy a^* is optimal for the agent if and only if, for almost all $t \in [0, T]$:

$$a_t^* \in \arg \max_{a_t} \theta_t a_t - g(a_t). \quad (57)$$

To prove this claim, consider another action policy (a_t) , adapted to the filtration induced by $(Z_s)_{s \leq t}$. Consider the value $W = v_T - \int_0^T g(a_t) dt$, so that the final utility for the agent under policy a is $u(W)$. Defining $L \equiv \int_0^T [\theta_t a_t - g(a_t) - \theta_t a_t^* + g(a_t^*)] dt$, it can be rewritten

$$W = v_0 + \int_0^T \theta_t (dr_t - a_t dt) - \int_0^T g(a_t^*) dt + L.$$

Suppose that (57) is not verified on the set τ of times with positive measure. Then, consider a policy a such that $\theta_t a_t - g(a_t) > \theta_t a_t^* - g(a_t^*)$ for $t \in \tau$, and $a_t = a_t^*$ on $[0, T] \setminus \tau$. We thus

have $L > 0$. Consider the agent's utility under policy a :

$$\begin{aligned}
U^a &= E^a \left[u \left(v_T - \int_0^T g(a_t) dt \right) \right] = E^a \left[u \left(v_0 + \int_0^T \theta_t (dr_t - a_t dt) - \int_0^T g(a_t^*) dt + L \right) \right] \\
&= E^a \left[u \left(v_0 + \int_0^T \theta_t \sigma_t dZ_t^a - \int_0^T g(a_t^*) dt + L \right) \right] \\
&> E^a \left[u \left(v_0 + \int_0^T \theta_t \sigma_t dZ_t^a - \int_0^T g(a_t^*) dt \right) \right] \text{ since } L > 0 \\
&= E^{a^*} \left[u \left(v_0 + \int_0^T \theta_t \sigma_t dZ_t^{a^*} - \int_0^T g(a_t^*) dt \right) \right] = E^{a^*} \left[u \left(v_T - \int_0^T g(a_t^*) dt \right) \right] = U^{a^*},
\end{aligned}$$

where U^{a^*} is the agent's utility under policy a^* . Hence, as $U^a > U^{a^*}$, the IC condition is violated. We conclude that condition (57) is necessary for the contract to satisfy the IC condition.

We next show that condition (57) is also sufficient to satisfy the IC condition. Indeed, consider any adapted policy a . Then, $L \leq 0$. So, the above reasoning shows that $U^a \leq U^{a^*}$. Policy a^* is at least as good as any alternative strategy a .

2) We show that cost-minimization entails $\theta_t = g'(a_t^*)$.

(57) implies $\theta_t = g'(a_t^*)$ if $a_t^* \in (\underline{a}, \bar{a})$, and $\theta_t \geq g'(a_t^*)$ if $a_t^* = \bar{a}$.

The case where $a_t^* \in (\underline{a}, \bar{a}) \forall t$ is straightforward. The IC contract must have the form:

$$v(c_T) = v_0 + \int_0^T g'(a_t^*) (dr_t - a_t^* dt) = \int_0^T g'(a_t^*) dr_t + K,$$

where $K = v_0 + \int_0^T g'(a_t^*) a_t^* dt$. Cost minimization entails the lowest possible v_0 .

The case where $a_t^* = \bar{a}$ for some t is more complex, since the IC constraint is only an inequality: $\theta_t \geq \theta_t^* \equiv g'(a_t^*)$. We must therefore prove this inequality binds. Consider

$$X = \int_0^T \theta_t^* \sigma_t dz_t, \quad Y = \int_0^T \theta_t \sigma_t dz_t.$$

By reshifting $u(x) \rightarrow u\left(x - \int_0^T g(a_t^*) dt\right)$ if necessary, we can assume $\int_0^T g(a_t^*) dt = 0$ to simplify notation.

We wish to show that a contract $v_T = Y + K_Y$, with $E[u(Y + K_Y)] \geq \underline{u}$, has a weakly greater expected cost than a contract $v = X + K_X$, with $E[u(X + K_X)] = \underline{u}$. Lemma 8 implies that $E[u(X + K_X)] \geq E[u(Y + K_X)]$, and so

$$E[u(Y + K_X)] \leq E[u(X + K_X)] = \underline{u} \leq [u(Y + K_Y)].$$

Thus, $K_X \leq K_Y$. Since v is increasing and concave, v^{-1} is convex and $-v^{-1}$ is concave. We

can therefore apply Lemma 8 to function $-v^{-1}$ to yield:

$$\mathbb{E} [v^{-1} (X + K_X)] \leq \mathbb{E} [v^{-1} (Y + K_X)] \leq \mathbb{E} [v^{-1} (Y + K_Y)],$$

where the second inequality follows from $K_X \leq K_Y$. Therefore, the expected cost of $v = X + K_X$ is weakly less than that of $Y + K_Y$, and so contract $v = X + K_X$ is cost-minimizing. More explicitly, that is the contract (22) with $K = K_X + \int_0^T g'(a_t^*) a_t^* dt$.

Proof of Proposition 2

The proof is by induction.

Proof of Proposition 2 for $T = 1$. We remove time subscripts and let $V(\hat{\eta}) = v(C(\hat{\eta}))$ denote the felicity received by the agent if he announces $\hat{\eta}$ and signal $A(\hat{\eta}) + \hat{\eta}$ is revealed.

If the agent reports η , the principal expects to see signal $\eta + A(\eta)$. Therefore, if the agent deviates to report $\hat{\eta} \neq \eta$, he must take action a such that $\eta + a = \hat{\eta} + A(\hat{\eta})$, i.e. $a = A(\hat{\eta}) + \hat{\eta} - \eta$. Hence, the truth-telling constraint is: $\forall \eta, \forall \hat{\eta}$,

$$V(\hat{\eta}) - g(A(\hat{\eta}) + \hat{\eta} - \eta) \leq V(\eta) - g(A(\eta)). \quad (58)$$

Defining

$$\psi(\eta) \equiv V(\eta) - g(A(\eta)),$$

the truth-telling constraint (58) can be rewritten,

$$g(A(\hat{\eta})) - g(A(\hat{\eta}) + \hat{\eta} - \eta) \leq \psi(\eta) - \psi(\hat{\eta}). \quad (59)$$

Rewriting this inequality interchanging η and $\hat{\eta}$ and combining with the original inequality (59) yields:

$$\forall \eta, \forall \hat{\eta} : g(A(\hat{\eta})) - g(A(\hat{\eta}) + \hat{\eta} - \eta) \leq \psi(\eta) - \psi(\hat{\eta}) \leq g(A(\eta) + \eta - \hat{\eta}) - g(A(\eta)). \quad (60)$$

Consider a point η where A is continuous and take $\hat{\eta} < \eta$. Dividing (60) by $\eta - \hat{\eta} > 0$ and taking the limit $\hat{\eta} \uparrow \eta$ yields $\psi'_{left}(\eta) = g'(A(\eta))$. Next, consider $\hat{\eta} > \eta$. Dividing (60) by $\eta - \hat{\eta} < 0$ and taking the limit $\hat{\eta} \downarrow \eta$ yields $\psi'_{right}(\eta) = g'(A(\eta))$. Hence,

$$\psi'(\eta) = g'(A(\eta)), \quad (61)$$

at all points η where A is continuous.

Equation (61) holds only almost everywhere, since we have only assumed that A is almost everywhere continuous. To complete the proof, we require a regularity argument about ψ (otherwise ψ might jump, for instance). We will show that ψ is absolutely continuous (see, e.g., Rudin (1987), p.145). Consider a compact subinterval I , and $\bar{a}_I = \sup \{A(\eta) + \eta - \hat{\eta} \mid \eta, \hat{\eta} \in I\}$, which

is finite because A is assumed to be bounded in any compact subinterval of η . Then, equation (60) implies:

$$|\psi(\eta) - \psi(\hat{\eta})| \leq \max \{|g(A(\hat{\eta})) - g(A(\hat{\eta}) + \hat{\eta} - \eta)|, g(A(\eta) + \eta - \hat{\eta}) - g(A(\eta))\} \leq |\eta - \hat{\eta}| (\sup g')_I.$$

This implies that ψ is absolutely continuous on I . Therefore, by the fundamental theorem of calculus for almost everywhere differentiable functions (Rudin (1987), p.148), we have that for any $\eta, \underline{\eta}$, $\psi(\eta) = \psi(\underline{\eta}) + \int_{\underline{\eta}}^{\eta} \psi'(x) dx$. From (61), $\psi(\eta) = \psi(\underline{\eta}) + \int_{\underline{\eta}}^{\eta} g'(A(x)) dx$, i.e.

$$V(\eta) = g(A(\eta)) + \int_{\underline{\eta}}^{\eta} g'(A(x)) dx + k \quad (62)$$

with $k = \psi(\underline{\eta})$. This concludes the derivation of the contract when $T = 1$.

“*Second-order conditions.*” We next show that the contract (62) does implement effort $A(\eta)$, iff $A(\eta) + \eta$ is nondecreasing: we have verified the first order condition, but we need to show that (58) holds given the proposed contract, that is, that $\Phi(\hat{\eta}) \equiv V(\hat{\eta}) - g(A(\hat{\eta}) + \hat{\eta} - \eta)$ has a maximum at η .

Proof that $A(\eta) + \eta$ nondecreasing is a sufficient condition for the contract to implement the action. First, we do this when $A(\eta)$ is a C^1 function. Then,

$$\begin{aligned} \Phi'(\hat{\eta}) &= V'(\hat{\eta}) - g'(A(\hat{\eta}) + \hat{\eta} - \eta)(A'(\hat{\eta}) + 1) \\ &= [g'(A(\hat{\eta})) - g'(A(\hat{\eta}) + \hat{\eta} - \eta)](A'(\hat{\eta}) + 1) \end{aligned}$$

As $A'(\hat{\eta}) + 1 \geq 0$ and g is convex, we have $\Phi'(\hat{\eta}) \geq 0$ for $\hat{\eta} \leq \eta$ and $\Phi'(\hat{\eta}) \leq 0$ for $\hat{\eta} \geq \eta$. That shows that $\Phi(\hat{\eta})$ is maximized at $\hat{\eta} = \eta$.

Second, in the case where A is not necessarily C^1 , we approximate the weakly increasing function $A(\eta) + \eta$ by a series of C^1 weakly increasing functions $A_n(\eta) + \eta$. (It is well-known that this is easy to do by convolution: take a random variable ε with bounded support and C^1 density f , and define $A_n(\eta) + \eta = E[A(\eta + \frac{\varepsilon}{n}) + \eta + \frac{\varepsilon}{n}] = \int (A(x) + x) f(n(x - \eta)) ndx$ which increasing in η by the first equality, and C^1 by the second.) Consider the associated contract $V_n \rightarrow V$. We have seen that $\eta \in \arg \max_{\hat{\eta}} V_n(\hat{\eta}) - g(A_n(\hat{\eta}) + \hat{\eta} - \eta)$, so in the limit, $\eta \in \arg \max_{\hat{\eta}} V(\hat{\eta}) - g(A(\hat{\eta}) + \hat{\eta} - \eta)$.

Proof that $A(\eta) + \eta$ nondecreasing is a necessary condition. Call $R(\eta) = A(\eta) + \eta$. Suppose by contradiction that there are two points $\eta < \eta'$ such that $R(\eta) > R(\eta')$. Those two points can be taken arbitrarily close (indeed, consider a large N , the points $\eta_i = \eta + (\eta' - \eta) i/N$, $i = 0 \dots N$; there must be an i such that $R(\eta_i) > R(\eta_{i+1})$, otherwise we would have $R(\eta) = R(\eta_0) \leq R(\eta_N) = R(\eta')$). As domain \mathcal{A} of actions is open, that implies that $A(\eta) + \eta - \eta' \in \mathcal{A}$. Applying (58) at point η and η' , we have:

$$V(\eta') - g(A(\eta') + \eta' - \eta) \leq V(\eta) - g(A(\eta)) \quad \text{and} \quad V(\eta) - g(A(\eta) + \eta - \eta') \leq V(\eta') - g(A(\eta')) \Rightarrow$$

$$g(A(\eta')) - g(A(\eta) + \eta - \eta') \leq V(\eta') - V(\eta) \leq g(A(\eta') + \eta' - \eta) - g(A(\eta))$$

Calling $y \equiv A(\eta) + \eta - \eta' < x \equiv A(\eta)$ and $h = A(\eta') + \eta' - A(\eta) - \eta$, this writes $g(y + h) - g(y) \leq g(x + h) - g(x)$, and we have a contradiction if g is strictly convex.

Proof that if Proposition 2 holds for T , it holds for $T + 1$. This part of the proof is as the proof of Theorem 1 in the main text. At $t = T + 1$, if the agent reports $\hat{\eta}_{T+1}$, he must take action $a = A(\hat{\eta}_{T+1}) + \hat{\eta}_{T+1} - \eta_{T+1}$ so that the signal $a + \eta_{T+1}$ is consistent with declaring $\hat{\eta}_{T+1}$. The IC constraint is therefore:

$$\eta_{T+1} \in \arg \max_{\hat{\eta}_{T+1}} V(\eta_1, \dots, \eta_T, \hat{\eta}_{T+1}) - g(A(\hat{\eta}_{T+1}) + \hat{\eta}_{T+1} - \eta_{T+1}) - \sum_{t=1}^T g(a_t^*). \quad (63)$$

Applying the result for $T = 1$, to induce $\hat{\eta}_{T+1} = \eta_{T+1}$, the contract must be of the form:

$$V(\eta_1, \dots, \eta_T, \hat{\eta}_{T+1}) = W_{T+1}(\hat{\eta}_{T+1}) + k(\eta_1, \dots, \eta_T), \quad (64)$$

where $W_{T+1}(\hat{\eta}_{T+1}) = g(A(\hat{\eta}_{T+1})) + \int_{\underline{\eta}}^{\hat{\eta}_{T+1}} g'(A(x)) dx$ and $k(\eta_1, \dots, \eta_T)$ is the ‘‘constant’’ viewed from period $T + 1$.

In turn, $k(\eta_1, \dots, \eta_T)$ must be chosen to implement $\hat{\eta}_t = \eta_t \forall t = 1 \dots T$, viewed from time 0, when the agent’s utility is:

$$E \left[u \left(k(\eta_1, \dots, \eta_T) + W_{T+1}(\hat{\eta}_{T+1}) - \sum_{t=1}^T g(a_t) \right) \right].$$

Defining

$$\hat{u}(x) = E[u(x + W_{T+1}(\tilde{\eta}_{T+1}))], \quad (65)$$

the principal’s problem is to implement $\hat{\eta} = \eta_t \forall t = 1 \dots T$, with a contract $k(\eta_1, \dots, \eta_T)$, given a utility function $E[\hat{u}(k(\eta_1, \dots, \eta_T) - \sum_{t=1}^T g(a_t))]$. Applying the result for T , we see that k must be:

$$k(\eta_1, \dots, \eta_T) = \sum_{t=1}^T g(A_t(\eta_t)) + \sum_{t=1}^T \int_{\underline{\eta}}^{\eta_t} g'(A_t(x)) dx + k_*$$

for some constant k_* . Combining this with (62), the only incentive compatible contract is:

$$V(\eta_1, \dots, \eta_T, \eta_{T+1}) = \sum_{t=1}^{T+1} g(A_t(\eta_t)) + \sum_{t=1}^{T+1} \int_{\underline{\eta}}^{\eta_t} g'(A_t(x)) dx + k_*.$$

The treatment of the second-order conditions ($A_t(\eta_t) + \eta_t$ nondecreasing) is as with $T = 1$.

Proof of Lemma 1

Step 1. It is easier to work in terms of $Q(\eta) = g'(A(\eta))$, the marginal cost of effort associated with plan $A(\eta)$. With a slight abuse of notation, define $C[Q]$ as the ex-

pected cost of implementing plan $Q = \{Q(\eta)\}$. From Proposition 2 with $T = 1$, $c(\eta, Q) = v^{-1}(g \circ (g')^{-1}(Q(\eta)) + \int_0^\eta Q(x) dx + K)$, where K is the solution of $E[u(\int_0^\eta Q(x) dx + K)] = \underline{u}$. Then, the expected cost is: $C[Q] = E[c(\eta, Q)]$.

We first establish that the contract cost $C[Q]$ is convex in the plan Q . Consider two plans Q^1 and Q^2 , $\theta_1 + \theta_2 = 1$ with $\theta_1, \theta_2 \in [0, 1]$, and the plan Q defined by $Q(\eta) = \theta_1 Q^1(\eta) + \theta_2 Q^2(\eta)$. Since u is concave,

$$E \left[u \left(\int_0^\eta Q(x) dx + \theta_1 K_1 + \theta_2 K_2 \right) \right] \geq \underline{u}$$

so the constant K associated with the new plan satisfies $K \leq \theta_1 K_1 + \theta_2 K_2$. This shows that the function $K[Q]$ is convex in Q . Since $g \circ (g')^{-1}$ and v^{-1} are convex, $C[Q] \leq \theta_1 C[Q_1] + \theta_2 C[Q_2]$, i.e., C is convex.

Step 2. Since C is convex, we have:²³

$$C[\bar{Q}] - C[Q] \leq \int \frac{\partial C[\bar{Q}]}{\partial Q(\eta)} (\bar{Q} - Q(\eta)) d\eta.$$

Furthermore, since g' is convex, $\bar{Q} - Q(\eta) \leq g''(\bar{a}) (\bar{a} - A(\eta))$. Defining $\lambda(\bar{a}, \eta) = \max\left(0, \frac{\partial C[\bar{A}]}{\partial A(\eta)}\right)$, we have $C[\bar{A}] - C[A] \leq \int \lambda(\bar{a}, \eta) (\bar{a} - a(\eta)) d\eta$.

C. 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.

²³We use partial derivatives such $\partial C[Q]/\partial Q(\eta)$. Their meaning is traditional and is as follows. Under weak conditions, $C[\cdot]$ is differentiable in Q , in the sense that there is a function $\xi(\eta)$ (unique up to sets of measure 0) such that, for any $\{B(\eta)\}$ such that $C[Q + hB]$ is well-defined for $h \geq 0$ close enough to 0, $\lim_{h \rightarrow 0} (C[Q + hB] - C[Q])/h = \int \xi(\eta) B(\eta) d\eta$. Then, we define $\partial C[Q]/\partial Q(\eta) = \xi(\eta)$. As C is convex, such a derivative exists almost everywhere (otherwise, for the purposes of the inequality, we can just take a subgradient.). As $Q(\eta) = g'(A(\eta))$, the chain rule gives the definition: $\partial C[A]/\partial A(\eta) \equiv \partial C[Q]/\partial Q(\eta) \cdot g''(A(\eta))$.

Proposition 3 (*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

$$L\theta = g'(\mathbf{a}^*), \quad (66)$$

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

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

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:

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

$$\forall \mathbf{r}, LK'(\mathbf{r}) = 0 \quad (68)$$

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

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:

$$V'(\mathbf{b}(\mathbf{a}^*) + \boldsymbol{\eta}) \mathbf{b}'(\mathbf{a}^*) - g'(\mathbf{a}^*) = 0, \quad (69)$$

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 (69) gives: $V(\mathbf{r}) = \theta \mathbf{r} + K(\mathbf{r})$, where $\theta \mathbf{r} = \sum_{i=1}^S \theta_i r_i$, 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 $\theta = (\theta_1, \theta_2) \in \mathbb{R}^2$, (66) becomes: $\theta_1 + \theta_2 = g'(a^*)$. For example, we can take $\theta_1 = \theta_2 = g'(a^*)/2$. Next, (68) 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 $\theta = (\theta_1, \theta_2) \in \mathbb{R}^2$, (66) becomes: $\theta_1 = g'(a^*)$. Next, (68) 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.

D. Extension to Optimal Contract With Noise-Dependent Actions

D.1. Affine Cost of Effort

While Theorem 2 shows that $A(\eta) = \bar{a}$ is optimal when Proposition 1 is satisfied, we now show that $A(\eta)$ can be exactly derived even if Theorem 2 does not hold, if the cost function is linear – i.e. $g(a) = \theta a$, where $\theta > 0$.²⁴ The optimal effort level is now interior. We use the benefit function $b(a, \eta) = S b_*(a, \eta)$ as in Section 2.2.

Proposition 4 (*Optimal contract with linear cost of effort*). *Let $g(a) = \theta a$, where $\theta > 0$. The following contract is optimal:*

$$c = v^{-1} (\theta r + K), \tag{70}$$

where K is a constant that makes the participation constraint bind ($\mathbb{E}[u(\theta \eta + K)] = \underline{u}$). For

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 $g(a)$.

each η , the optimal effort $A(\eta)$ is determined by the following pointwise maximization:

$$A(\eta) \in \arg \max_{a \leq \bar{a}} S b_*(a, \eta) - v^{-1}(\theta(a + \eta) + K). \quad (71)$$

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 $V(\eta) = g(A(\eta)) + \int_{\eta}^{\infty} \theta dx + K = \theta(A(\eta) + \eta) + K$. 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:

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

which is solved by pointwise maximization, as in (71). ■

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 (71) 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.

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 4 therefore assumes that indeterminacies are resolved by the agent following the principal's recommended action, $A(\eta)$.

D.2. Conditions for Optimality of High Effort

Section 2.2 showed that the condition in Theorem 2,

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

required for high effort to be optimal, 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 $L(\eta) = \int_{\underline{\eta}}^{\eta} g'(a(x)) dx$, $\underline{\eta}$ 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 5 *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:*

$$\begin{aligned} \frac{\partial C[A]}{\partial a(\eta)} = & \frac{g'(a(\eta))}{v'(c(\eta))} f(\eta) + \\ & g''(a(\eta)) \left\{ E \left[\frac{1}{v'(c(\tilde{\eta}))} 1_{\tilde{\eta} > \eta} \right] - E \left[\frac{1}{v'(c(\tilde{\eta}))} \right] \frac{E[u'(L(\tilde{\eta}) + K) 1_{\tilde{\eta} > \eta}]}{E[u'(L(\tilde{\eta}) + K)]} \right\}. \end{aligned} \quad (72)$$

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

The first term in (72), $\frac{g'(a(\eta))}{v'(c(\eta))} f(\eta)$, 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 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 (72) 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

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 (72).

distributions; in addition, we can establish bounds on it to help verify whether Proposition 1 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

$$\frac{\partial C [A]}{\partial a(\eta)} \leq \Lambda(\bar{a}, \eta) \equiv \frac{g'(\bar{a})f(\eta) + g''(\bar{a})\bar{F}(\eta)}{v'(v^{-1}(u^{-1}(\underline{u}) + g(\bar{a}) + (\bar{\eta} - \underline{\eta})g'(\bar{a})))}. \quad (73)$$

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 (73) 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

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

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 (72) 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 (73).

In the model of section 2.2, using the notations in (28), equation (73) implies: $\lambda(\bar{a}, \eta) \leq \frac{\Lambda(\bar{a}, \eta)}{f(\eta)} \leq \Lambda(\bar{a})$. Hence, by Theorem 2, highest effort is optimal if for all η , $S \partial_1 b_*(a, \eta) f(\eta) \geq f(\eta) \Lambda(\bar{a})$, which is verified if $S \geq S_* \equiv \frac{\Lambda(\bar{a})}{\inf_{\eta} \partial_1 b_*(\bar{a}, \eta)}$.

D.3. Illustrations for Proposition 3

We now provide explicit conditions to verify the optimality of high effort in the two examples in Appendix 2.3. We also analyze a third example.

Example 1. 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 (72) 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 3, a^{**} will be the optimum if

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

for all η .

Example 2. 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. Theorem 3 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 (73) 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 Theorem 3, 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 3: HM framework. Consider $v(x) = x$, $g(a) = \frac{1}{2}Ga^2$, $u(x) = -e^{-\gamma x}$ with $G, \gamma > 0$, and $\eta \sim N(0, \sigma^2)$ as in HM. The cost of the contract is $C[a] = \underline{c} + g(a) + \gamma g'(a)^2 \sigma^2/2$, and the same comparative statics with respect to G , σ and γ hold.

HM have not only a constant target action, but also an additive effect of effort. We can obtain this result with $b(a, \eta, \bar{a}) = a + (a - \bar{a}) \beta(\bar{a}, \eta)$ as in Examples 1 and 2. The key nuance in obtaining the HM result is reconciling the linear marginal benefit of effort required for an additive effect, with the large marginal benefit of effort required for high effort to be optimal. The two-stage game resolves this tension because the marginal benefit of effort is moderate in the first stage and very high in the second stage.

Under this formulation for $b(a, \eta, \bar{a})$, the cost of the contract implementing $a = \bar{a}$ is

$$C[a] = \underline{c} + \frac{1}{2}Ga^2 + \frac{\gamma}{2}G^2a^2\sigma^2$$

and the principal maximizes

$$a - \underline{c} - \frac{1}{2}Ga^2 - \frac{\gamma}{2}G^2a^2\sigma^2$$

which yields the result

$$a = 1/[G(1 + G\gamma\sigma^2)].$$

This is the same optimal contract as in HM, but without requiring continuous time.

In terms of verifying Theorem 3, we have

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

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 Theorem 3, a^{**} will be the optimum if

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

for all η .

E. 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 Remark 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, high 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 6 (*Contract with firing, $T = 2$*). *Under the conditions of Remark 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 Remark 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:

$$c = v^{-1} \left(\sum_{t=1}^{\tau} g'(a_t^*) r_t + K_{\tau} \right) \quad (74)$$

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 6, but with K_1 and K_2 are simply set high enough such that the agent always receives at least his reservation utility.

F. A Microfoundation for the Principal’s Objective

We offer a microfoundation for the principal’s objective function (26). Suppose that the agent can take two actions, a “fundamental” action $a^F \in (\underline{a}, \bar{a}]$ and a manipulative action $m \geq 0$. Firm value is a function of a^F only, i.e. the benefit function is $b(a^F, \eta)$. The signal is increasing in both actions: $r = a^F + m + \eta$. The agent’s utility is $v(c) - [g^F(a) + G(m)]$, where g, G are increasing and convex, $G(0) = 0$, and $G'(0) \geq g'(\bar{a})$. The final assumption means that manipulation is costlier than fundamental effort.

We define $a = a^F + m$ and the cost function $g(a) = \min_{a^F, M} \{g^F(a) + G(m) \mid a^F + m = a\}$, so that $g(a) = g^F(a)$ for $a \in (\underline{a}, \bar{a}]$ and $g(a) = g^F(a) + g(m - a)$ for $a \geq \bar{a}$, which is increasing and convex. Then, firm value can be written $b(\min(a, \bar{a}), \tilde{\eta})$, as in equation (26).

This framework is consistent with rational expectations. Suppose $b(a^F, \eta) = e^{a^F + \eta}$. After observing the signal r , the market forms its expectation P_1 of the firm value $b(a^F, \eta)$. The incentive contract described in Proposition 2 implements $a \leq \bar{a}$, so the agent will not engage in manipulation. Therefore, the rational expectations price is $P_1 = e^r$.

In more technical terms, consider the game in which the agent takes action a and the market sets price P_1 after observing signal r . It is a Bayesian Nash equilibrium for the agent to choose $A(\eta)$ and for the market to set price $P_1 = e^r$.

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