Tractability and Detail-Neutrality in Incentive Contracting*

Alex Edmans
Wharton

Xavier Gabaix
NYU Stern and NBER

March 22, 2009

Abstract

This paper identifies a broad class of situations in which the contract is both attainable in closed form and "detail-neutral". The contract's functional form is independent of the noise distribution and reservation utility; moreover, when the cost of effort is pecuniary, the contract is linear in output regardless of the agent's utility function. Our contract holds in both continuous time and a discrete-time, multi-period setting where action follows noise in each period. The tractable contracts of Holmstrom and Milgrom (1987) can thus be achieved in settings that do not require exponential utility, Gaussian noise or continuous time. Our results also suggest that incentive schemes need not depend on complex details of the particular setting, a number of which (e.g. agent's risk aversion) are difficult for the principal to observe. The proof techniques use the notion of relative dispersion and subdifferentials to avoid relying on the first-order approach, and may be of methodological interest.

KEYWORDS: Contract theory, executive compensation, incentives, principal-agent problem, dispersive order, subderivative. (JEL: D2, D3, G34, J3)

---

*aedmans@wharton.upenn.edu, xgabaix@stern.nyu.edu. For helpful comments, we thank three anonymous referees, the Editor (Stephen Morris), Andy Abel, Heski Bar-Isaac, Patrick Cheridito, Peter DeMarzo, Ingolf Dittmann, Florian Ederer, Chris Evans, Itay Goldstein, Gary Gorton, Ernst Maug, Holger Mueller, David Pearce, Canice Prendergast, Michael Roberts, Nick Souleles, Rob Tumarkin, Tomasz Sadzik, Yuliy Sannikov, and seminar participants at AEA, Five Star, Chicago GSB, Harvard-MIT Organizational Economics, Minneapolis Fed, Northwestern, NYU, Richmond Fed, Toulouse, Wisconsin, and Wharton. We thank Qi Liu for excellent research assistance. AE is grateful for the hospitality of the NYU Stern School of Business, where part of this research was carried out. XG thanks the NSF for financial support.

1 Introduction

The principal-agent problem is central to many economic settings, such as employment contracts, insurance and regulation. A vast literature analyzing this problem has found that it is typically difficult to solve: even in simple settings, the optimal contract can be highly complex (see, e.g., Grossman and Hart (1983)). The first-order approach is often invalid, requiring the use of more intricate techniques. Even if an optimal contract can be derived, it may not be attainable in closed form, which reduces tractability – a particularly important feature in applied theory models. In addition, the contract is typically contingent upon many specific features of the environment, a number of which (such as the agent's risk aversion) are difficult to observe and thus use to guide the contract design in practice. Even those parameters that can be observed do not appear to affect real-life contracts as much as existing theories predict.

Against this backdrop, Holmstrom and Milgrom (1987, "HM") made a major breakthrough by showing that the optimal contract is linear in profits under certain conditions. Their result has since been widely used by applied theorists to justify the focus on linear schemes, which leads to substantial tractability. However, HM emphasized that their result only holds under exponential utility, continuous time, Gaussian noise, and a pecuniary cost of effort. In certain settings, the modeler may wish to use discrete time or binary noise for clarity, or decreasing absolute risk aversion for empirical consistency.

Can tractable contracts be achieved in broader settings? When allowing for alternative utility functions or noise distributions, do these details start to affect the optimal contract? What factors do and do not matter for the incentive scheme? These questions are the focus of our paper. We first consider a discrete-time, multiperiod model where, in each period, the agent first observes noise and then exerts effort, before observing the noise in the next period; he consumes only in the final period. We solve for the cheapest contract that implements a given, but possibly time-varying, path of target effort levels. The optimal incentive scheme is both tractable (i.e. attainable in closed form) and "detail-neutral:" its functional form is independent of the noise distribution and the agent's reservation utility, and depends only on how the agent trades off the benefits of cash against the cost of providing effort.[1] The irrelevance of the noise distribution occurs even though each action, except the final one, is followed by noise, and so he faces uncertainty when deciding his effort level. Using recent advances in continuous-time contracting (Sannikov (2008)),

[1] For brevity, we call such a contract "detail-neutral." This term emphasizes that certain details of the contracting situation do not matter for the functional form of the contract (whereas they matter in earlier theories). It is not meant to imply that the functional form is independent of all parameters. Note that our notion of detail-neutral contracts is a separate concept from Wilson's (1987) detail-free auction mechanisms.

we show that the contract has the same functional form in a continuous-time model where noise and effort occur simultaneously.

In addition to noise and reservation utility, the optimal contract is also independent of the agent's utility function in two cases. First, if the cost of effort is pecuniary as in HM (i.e. can be expressed as a subtraction to cash pay), the incentive scheme is linear in output for any utility function, even if the cost of effort is itself non-linear. Second, if the agent's preferences are multiplicative in cash and effort, the optimal contract is independent of utility and log-linear, i.e. the percentage change in pay is linear in output. An application to CEO incentives demonstrates the implications that flow from a tractable contract structure. For CEOs, the appropriate output measure is the percentage stock return, and multiplicative preferences are theoretically motivated by Edmans, Gabaix and Landier (2008). The optimal contract thus sets the required percentage change in pay for a percentage change in firm value, i.e. the elasticity of pay with respect to firm value. This analysis provides a theoretical justification for using elasticities to measure incentives, a metric previously advocated by Murphy (1999) on empirical grounds.

We allow the target effort path to depend on the noise realizations. The optimal contract now depends on messages sent by the agent to the principal regarding the noise, since the "state of nature" may affect the productivity of effort. However, it remains tractable and detail-neutral, for a given "action function" that links the observed noise to the principal's recommended effort level. We finally solve for the optimal action function chosen by the principal. In classical agency models, the action chosen by the principal is the result of a trade-off between the benefits of effort (which are increasing in firm size) and its costs (direct disutility plus the risk imposed by incentives, which are of similar order of magnitude to the CEO's wage). We show that, if the output under the agent's control is sufficiently large compared his salary (e.g. the agent is a CEO who affects total firm value), these trade-off considerations disappear: the benefits of effort swamp the costs. Thus, maximum effort is optimal in each period, regardless of the noise outcome. By contrast, if output is small, maximum effort may not be optimal for some noise realizations. We show that the optimal action function can still be solved for if the cost function is affine.

The "maximum effort principle"[2], when applicable, significantly increases tractability, since it removes the need to solve the trade-off required to derive the optimal effort level when it is interior. Indeed, jointly deriving the optimal effort level and the efficient contract that implements it can be highly complex. Thus, papers that analyze the second (implementation) problem typically assume a fixed target effort level (e.g. Grossman and

[2] We allow for the agent to exert effort that does not benefit the principal. The "maximum effort principle" refers to the maximum *productive* effort that the agent can undertake to benefit the principal.

Hart (1983), Dittmann and Maug (2007) and Dittmann, Maug and Spalt (2008)). Our result rationalizes this approach: if maximum effort is always optimal, the first problem has a simple solution – there is no trade-off to be simultaneously tackled and the analysis can focus on the implementing contract.

In sum, for a given target effort level, the optimal implementation is detail-neutral. Moreover, if output is sufficiently large, the optimal action itself does not depend on model parameters, and so the overall contract is detail-neutral. All of the above results are derived under a general contracting framework, where the contract may depend on messages sent by the agent to the principal, and also be stochastic.

Our analytical framework yields a number of implications. First, it shows that tractable contracts can be derived even without assuming exponential utility, Gaussian noise and continuous time. This result may be of use for future contracting models, as it shows that tractability may be achieved in quite broad settings. For example, certain models may require decreasing relative risk aversion for empirical consistency and/or discrete time for clarity. Second, it demonstrates what details of the environment do and do not matter for the optimal contract. The functional form depends only on how the agent trades off cash against effort and not the noise distribution or reservation utility, and is independent of utility with a pecuniary cost of effort or multiplicative preferences. This detail-neutrality contrasts with many classical principal-agent models (e.g. Grossman and Hart (1983)), where the optimal contract is contingent upon many specific features of the contracting situation. This poses practical difficulties, as some of the important determinants are difficult to observe, such as the noise distribution and agent's utility function. Our results suggest that the contract is robust to such parametric uncertainty. Furthermore, even those parameters that can be observed do not appear to affect real-life contracts: for example, Prendergast's (2002) review of the evidence finds that incentives show little correlation with risk. Our model offers a simple potential explanation for why contracts typically do not depend on as many details of the contracting situation as a reading of the extant literature would suggest – these details in fact do not matter.

We achieve simple contracts in other settings than HM due to a different modeling setup. HM use exponential utility to eliminate "wealth effects" of prior period outcomes on the current period decision, thus removing the intertemporal link between periods and allowing the multiperiod problem to collapse into a succession of identical static problems. The removal of wealth effects also leads to independence of the reservation wage. By contrast, we achieve tractability by modeling the noise before the action in each period, as in theories in which the agent observes a "state of nature" before taking his action, or total output before deciding how much cash flow to divert.[3] This assumption has little effect on

[3]This timing assumption cannot be made within the HM framework since HM model effort as the

the economics of the problem since the agent is still exposed to risk in every period (except the final one), but is a technical device that allows the multiperiod model to be solved by backward induction, so that it becomes a succession of single-period problems. A potential intertemporal link remains since high past outcomes mean that the agent already expects high consumption and thus may have a lower current incentive to exert effort. This issue is present in the Mirrlees (1974) contract if the agent can observe past outcomes. (A high reservation wage has the same effect.) The optimal contract must address these issues: if the utility function is concave, the contract must be convex so that, at high levels of consumption, the agent is awarded a greater number of dollars for exerting effort, to offset the lower marginal utility of each additional dollar. If the cost of effort is in monetary terms, high past outcomes decrease the benefits of cash and the cost of effort equally, and so incentives are preserved even with a linear contract.

In addition to its results, the paper's proofs import and extend some mathematical techniques that are relatively rare in economic theory and may be of use in future models. We employ use the subderivative, a generalization of the derivative that allows for quasi first-order conditions even if the objective function is not everywhere differentiable. This concept is related to Krishna and Maenner's (2001) use of the subgradient, although the applications are quite different. These notions also allow us to avoid the first-order approach, and so may be useful for future models where sufficient conditions for the first-order approach cannot be verified.[4] We also use the notion of "relative dispersion" for random variables to prove that the incentive compatibility constraints bind, i.e. the principal imposes the minimum incentive slope that induces the target effort level. We show that the binding contract is less dispersed than alternative solutions, constituting efficient risk sharing. A similar argument rules out stochastic contracts, where the payout is a random function of output.[5] We extend a result from Landsberger and Meilijson (1994), who use relative dispersion in another economic setting.

This paper builds on a rich literature on the principal-agent problem. Grossman and Hart (1983) demonstrate how the problem can be solved in discrete time using a dynamic programming methodology that avoids the need for the first-order approach. HM show

selection of probabilities, so noise inevitably follows the action. We thus depart from the framework and model effort as an increment to mean output, so that the noise/state of nature can be realized first.

[4]See Rogerson (1985) for sufficient conditions for the first-order approach to be valid under a single signal, and Jewitt (1988) for situations in which the principal can observe multiple signals. Schaettler and Sung (1993) derive sufficient conditions for the first-order approach to be valid in a large class of principal-agent problems, of which HM is a special case.

[5]With separable utility, it is straightforward to show that the constraints bind: the principal should offer the least risky contract that achieves incentive compatibility. However, with non-separable utility, introducing additional randomization by giving the agent a riskier contract than necessary may be desirable (Arnott and Stiglitz (1988)) – an example of the theory of second best. We use the concept of relative dispersion to prove that constraints bind.

that optimal contracts are linear in profits in continuous time (where noise is automatically Gaussian) if the agent has exponential utility and controls only the drift of the process; they show that this result does not hold in discrete time. A number of papers have extended their result to more general settings, although all continue to require exponential utility. In Sung (1995) and Ou-Yang (2003), the agent also controls the diffusion of the process in continuous time. Hellwig and Schmidt (2002) show that linearity can be achieved in discrete time, under the additional assumptions that the agent can destroy profits before reporting them to the principal, and that the principal can only observe output in the final period. Our multiperiod model yields linear contracts while allowing the principal to observe signals at each interim stage. Mueller (2000) shows that linear contracts are not optimal in HM if the agent can only change the drift at discrete points, even if these points are numerous and so the model closely approximates continuous time. In a different setting from HM, where the agent can falsify the level of output, Lacker and Weinberg (1989) also identify a simple class of situations in which linear contracts obtain. Their core result is similar to a specific case of our Theorem 1, with a linear felicity function and a single period.

A number of other papers investigate the parameter dependence of optimal contracts. DeMarzo and Fishman (2007) consider a discrete-time, dynamic model where the agent can divert cash flows. They show that the optimal contract can be implemented using the standard securities of equity, long-term debt and a credit line; under certain conditions, the terms of debt and the credit line are independent of the severity of the agency problem. The agent is risk-neutral in their setting; here, we study the impact of the utility function and noise distribution. DeMarzo and Sannikov (2006) show that the model is particularly tractable in continuous time, where the incentive scheme can be solved as a differential equation. Wang (2007) derives the optimal contract under uncertainty and finds the limit of this contract as uncertainty diminishes. The limit contract depends on the agent's risk aversion and the characteristics of the risk environment.

This paper proceeds as follows. In Section 2 we derive tractable and detail-neutral contracts in both discrete and continuous time, as well as considering a specific application to CEO compensation. While this section holds the target effort level fixed, Section 3 allows it to depend on the noise realization and derives conditions under which the maximum productive effort level is optimal for all noise outcomes. Section 4 concludes. The Appendix contains proofs and other additional materials; further peripheral material is in the Online Appendix.

2 The Core Model

2.1 Discrete Time

We consider a T-period model; its key parameters are summarized in Table 1. In each period t, the agent observes noise η_t, takes an unobservable action a_t, and then observes the noise in period $t+1$. The action a_t is broadly defined to encompass any decision that benefits output but is personally costly to the principal. The main interpretation is effort, but it can also refer to rent extraction: low a_t reflects cash flow diversion or the pursuit of private benefits. We assume that noises $\eta_1, ..., \eta_T$ are independent with open interval support $(\underline{\eta_t}, \overline{\eta_t})$, where the bounds may be infinite, and that $\eta_2, ..., \eta_t$ have log-concave densities.[6] We require no other distributional assumption for η_t; in particular, it need not be Gaussian. The action space $\mathcal{A}$ has interval support, bounded below and above by $\underline{a}$ and $\overline{a}$. (We allow for both open and closed action sets and for the bounds to be infinite.) After the action is taken, a verifiable signal

$$r_t = a_t + \eta_t. \tag{1}$$

is publicly observed at the end of each period t.

Insert Table 1 about here

Our assumption that η_t precedes a_t is featured in models in which the agent observes a "state of nature" before taking his action (e.g. Harris and Raviv (1979), Laffont and Tirole (1986) and Baker (1992))[7] and cash flow diversion models where the agent observes total output before choosing how much to divert (e.g. DeMarzo and Sannikov (2006), DeMarzo and Fishman (2007).) Note that this timing assumption does not make the agent immune to risk – in every period, except the final one, his action is followed by noise. Even in a one-period model, the agent bears risk as the noise is unknown when he signs the contract. In Section 2.2 we show that the contract has the same functional form in continuous time, where η and a are simultaneous.

In period T, the principal pays the agent cash of c.[8] The agent's utility function is

---

[6] A random variable is log-concave if it has a density with respect to the Lebesgue measure, and the log of this density is a concave function. Many standard density functions are log-concave, in particular the Gaussian, uniform, exponential, Laplace, Dirichlet, Weibull, and beta distributions (see, e.g., Caplin and Nalebuff (1991)). On the other hand, most fat-tailed distributions are not log-concave, such as the Pareto distribution.

[7] In such papers, the optimal action typically depends on the state of nature. We allow for such dependence in Section 3.1.

[8] If the agent quits before time T, he receives a very low wage $\underline{c}$.

$$\mathrm{E}\left[u\left(v(c)-\sum_{t=1}^{T} g(a_t)\right)\right]. \quad (2)$$

g represents the cost of effort, which is increasing and weakly convex. u is the utility function and v is the felicity[9] function which denotes the agent's utility from cash; both are increasing and weakly concave. g, u and v are all twice continuously differentiable. We specify functions for both utility and felicity to maximize the generality of the setup. For example, the utility function $\frac{\left(ce^{-g(a)}\right)^{\frac{1}{\gamma}}}{1-\gamma}$ is commonly used in macroeconomics and features a non-linear u and v. $u(x) = x$ denotes additively separable preferences; $v(c) = \ln c$ generates multiplicative preferences. If $v(c) = c$, the cost of effort is expressed as a subtraction to cash pay. This is appropriate if effort represents an opportunity cost of foregoing an alternative income-generating activity (e.g. outside consulting), or involves a financial expenditure. Note that even if the cost of effort is pecuniary, it remains a general, possibly non-linear function $g(a_t)$. HM assume $v(c) = c$ and $u(x) = -e^{-\gamma x}$, i.e. a pecuniary cost of effort and exponential utility.

The only assumption that we make for the utility function u is that it exhibits nonincreasing absolute risk aversion (NIARA), i.e. $-u''(x)/u'(x)$ is nonincreasing in x. Many commonly used utility functions (e.g. constant absolute risk aversion $u(x) = -e^{-\gamma x}$ and constant relative risk aversion $u(x) = x^{1-\gamma}/(1-\gamma)$, $\gamma > 0$) exhibit NIARA. This assumption turns out to be sufficient to rule out randomized contracts.

The agent's reservation utility is given by $\underline{u} \in \operatorname{Im} u$, where $\operatorname{Im} u$ is the image of u, i.e. the range of values taken by u. We also assume that $\operatorname{Im} v = \mathbb{R}$ so that we can apply the v^{-1} function to any real number.[10] We impose no restrictions on the contracting space available to the principal, so the contract $\widetilde{c}(\cdot)$ can be stochastic, nonlinear in the signals r_t, and depend on messages M_t sent by the agent. By the revelation principle, we can assume that the the space of messages M_t is $\mathbb{R}$ and that the principal wishes to induce truth-telling by the agent. The full timing is as follows:

1. The principal proposes a (possibly stochastic) contract $\widetilde{c}(r_1, ..., r_T, M_1, ..., M_T)$
2. The agent agrees to the contract or receives his reservation utility $\underline{u}$.
3. The agent observes noise η_1, then sends the principal a message M_1, then exerts effort a_1.

[9] We note that the term "felicity" is typically used to denote one-period utility in an intertemporal model. We use it in a non-standard manner here to distinguish it from the utility function u.

[10] This assumption could be weakened. With K defined as in Theorem 1, it is sufficient to assume that there exists a value of K which makes the participation constraint bind, and a "threat consumption" which deters the agent from exerting very low effort, i.e. $\inf_c v(c) - \inf_{a_t} \sum_t g(a_t) \leq \sum_t g'(a^*)\left(\underline{\eta_t} + a_t^*\right) + K$.

4. The signal $r_1 = \eta_1 + a_1$ is publicly observed.

5. Steps (3)-(4) are repeated for $t = 2, ..., T$.

6. The principal pays the agent $\widetilde{c}(r_1, ..., r_T, M_1, ..., M_T)$.

As in Grossman and Hart (1983), in this section we fix the path of effort levels that the principal wants to implement at $(a_t^*)_{t=1,..,T}$, where $a_t^* > \underline{a}$ and a_t^* is allowed to be time-varying.[11] An admissible contract gives the agent an expected utility of at least $\underline{u}$ and induces him to take path (a_t^*) and truthfully report noises $(\eta_t)_{t=1,..,T}$. Since the principal is risk-neutral, the optimal contract is the admissible contract with the lowest expected cost $\mathrm{E}[\widetilde{c}]$.

We now formally define the principal's program. Let $\mathcal{F}_t$ be the filtration induced by $(\eta_1, ..., \eta_t)$, the noise revealed up to time t. The agent's policy is $(a, M) = (a_1, ..., a_T, M_1, ..., M_T)$, where a_t and M_t are $\mathcal{F}_t-$measurable. a_t is the effort taken by the agent if noise $(\eta_1, ..., \eta_t)$ has been realized, and M_t is a message sent by the agent upon observing $(\eta_1, ..., \eta_t)$. Let S denote the space of such policies, and $\Delta(S)$ the set of randomized policies. Define $(a^*, M^*) = (a_1^*, ..., a_T^*, M_1^*, ..., M_T^*)$ the policy of exerting effort a_t^* at time t, and sending the truthful message $M_t^*(\eta_1, ..., \eta_t) = \eta_t$. The program is given below:

Program 1 *The principal chooses a contract $\widetilde{c}(r_1, ..., r_T, M_1, ..., M_T)$ and a $\mathcal{F}_t-$measurable message policy $(M_t^*)_{t=1...,T}$, that minimizes expected cost:*

$$\min_{\widetilde{c}(\cdot)} \; \mathrm{E}\left[\widetilde{c}\left(a_1^* + \eta_1, ..., a_T^* + \eta_T, M_1^*, ..., M_T^*\right)\right], \tag{3}$$

subject to the following constraints:

$$\text{IC: } (a_t^*, M_t^*)_{t=1...T} \in \arg\max_{(a,M)\in\Delta(S)} \mathrm{E}\left[u\left(v\left(\widetilde{c}\left(a_1 + \eta_1, ..., a_T + \eta_T, M_1, ..., M_T\right)\right) - \sum_{s=1}^{T} g(a_s)\right)\right] \tag{4}$$

$$\text{IR: } \mathrm{E}\left[u\left(v\left(\widetilde{c}(\cdot)\right) - \sum_{t=1}^{T} g(a_t^*)\right)\right] \geq \underline{u}. \tag{5}$$

In particular, if the analysis is restricted to message-free contracts, (4) implies that the time-t action a_t^* is given by:

$$\forall \eta_1, ..., \eta_t, \; a_t^* \in \arg\max_{a_t} \mathrm{E}\left[u\left(v\left(\widetilde{c}\left(a_1^* + \eta_1, ..., a_t + \eta_t, ..., a_T^* + \eta_T\right)\right) - g(a_t) - \sum_{s=1, s\neq t}^{T} g(a_s^*)\right) \mid \eta_1, ..., \eta_t\right] \tag{6}$$

[11] If $a_t^* = \underline{a}$, then a flat wage induces the optimal action.

Theorem 1 below describes our solution to Program 1.[12]

Theorem 1 *(Optimal contract, discrete time). The following contract is optimal. The agent is paid*

$$c = v^{-1}\left(\sum_{t=1}^{T} g'\left(a_t^*\right) r_t + K\right), \tag{7}$$

where K is a constant that makes the participation constraint bind ($\mathrm{E}\left[u\left(\begin{array}{c}\sum_t g'\left(a_t^*\right) r_t + \\ K - \sum_t g\left(a_t^*\right)\end{array}\right)\right] = \underline{u}$). The functional form (7) is independent of the utility function u, the reservation utility $\underline{u}$, and the distribution of the noise η. These parameters affect only the scalar K. The optimal contract is deterministic and does not require messages.

In particular, if the target action is time-independent ($a_t^* = a^* \ \forall\ t$), the contract

$$c = v^{-1}\left(g'\left(a^*\right) r + K\right) \tag{8}$$

is optimal, where $r = \sum_{t=1}^{T} r_t$ is the total signal.

Proof. (Heuristic). The Appendix presents a rigorous proof that rules out stochastic contracts and messages, and does not assume that the contract is differentiable. Here, we give a heuristic proof by induction on T that conveys the essence of the result for deterministic message-free contracts, using first-order conditions and assuming $a_t^* < \overline{a}$. We commence with $T = 1$. Since η_1 is known, we can remove the expectations operator from the incentive compatibility condition (6). Since u is an increasing function, it also drops out to yield:

$$a_1^* \in \arg\max_{a_1} v\left(c\left(a_1 + \eta_1\right)\right) - g\left(a_1\right). \tag{9}$$

The first-order condition is:

$$v'\left(c\left(a_1^* + \eta_1\right)\right) c'\left(a_1^* + \eta_1\right) - g'\left(a_1^*\right) = 0.$$

Therefore, for all r_1,

$$v'\left(c\left(r_1\right)\right) c'\left(r_1\right) = g'\left(a_1^*\right),$$

which integrates over η_1 to

$$v\left(c\left(r_1\right)\right) = g'\left(a_1^*\right) r_1 + K \tag{10}$$

[12] Theorem 1 characterizes a contract that is optimal, i.e. solves Program 1. Strictly speaking, there exist other optimal contracts which pay the same as (7) on the equilibrium path, but take different values for returns that are not observed on the equilibrium path.

for some constant K. Contract (10) must hold for all r_1 that occurs with non-zero probability, i.e. for $r_1 \in \left(a_1^* + \underline{\eta}_1, a_1^* + \overline{\eta}_1\right)$.

Proceeding by induction, we now show that, if the result holds for T, it also holds for $T+1$. Let $V(r_1, ..., r_{T+1}) \equiv v(c(r_1, ..., r_{T+1}))$ denote the indirect felicity function, i.e. the contract in terms of felicity rather than cash. At $t = T+1$, the incentive compatibility condition is:

$$a_{T+1}^* \in \arg\max_{a_{T+1}} V(r_1, ..., r_T, \eta_{T+1} + a_{T+1}) - g(a_{T+1}) - \sum_{t=1}^{T} g(a_t^*). \tag{11}$$

Applying the result for $T = 1$, to induce a_{T+1}^* at $T+1$, the contract must be of the form:

$$V(r_1, ..., r_T, r_{T+1}) = g'\left(a_{T+1}^*\right) r_{T+1} + k(r_1, ..., r_T), \tag{12}$$

where the integration "constant" now depends on the past signals, i.e. $k(r_1, ..., r_T)$. In turn, $k(r_1, ..., r_T)$ must be chosen to implement $a_1^*, ..., a_T^*$ viewed from $t = 0$, when the agent's utility is:

$$E\left[u\left(k(r_1, ..., r_T) + g'\left(a_{T+1}^*\right) r_{T+1} - g\left(a_{T+1}^*\right) - \sum_{t=1}^{T} g(a_t)\right)\right].$$

Defining

$$\widehat{u}(x) = \mathrm{E}\left[u\left(x + g'\left(a_{T+1}^*\right) r_{T+1} - g\left(a_{T+1}^*\right)\right)\right], \tag{13}$$

the principal's problem is to implement $a_1^*, ..., a_T^*$ with a contract $k(r_1, ..., r_T)$, given a utility function

$$E\left[\widehat{u}\left(k(r_1, ..., r_T) - \sum_{t=1}^{T} g(a_t)\right)\right].$$

Applying the result for T, the contract must have the form $k(r_1, ..., r_T) = \sum_{t=1}^{T} g'(a_t^*) r_t + K$ for some constant K. Combining this with (10), an incentive compatible contract must satisfy:

$$V(r_1, ..., r_T, r_{T+1}) = \sum_{t=1}^{T+1} g'(a_t^*) r_t + K. \tag{14}$$

for (r_t) that occurs with non-zero probability (i.e. $(r_1, ..., r_T) \in \prod_{t=1}^{T}\left(a_t^* + \underline{\eta}_t, a_t^* + \overline{\eta}_t\right)$. The associated pay is $c = v^{-1}\left(\sum_{t=1}^{T+1} g'(a_t^*) r_t + K\right)$, as in (7). Conversely, any contract that satisfies (14) is incentive compatible. ■

The main applications of Theorem 1 are likely to be for $T = 1$ or for a constant a_t^*; Section 3.2 derives conditions under which the maximum productive effort level is optimal for all t. In such cases, the contract is particularly simple and only depends on the total signal, as shown in (8).

In addition to deriving the incentive scheme in closed form (for any T and (a_t^*)), Theorem 1 also clarifies the parameters that do and do not matter for the contract's functional form. It depends only on the felicity function v and the cost of effort g, i.e. how the agent trades off the benefits of cash against the costs of providing effort, and is independent of the utility function u, the reservation utility $\underline{u}$, and the distribution of the noise η. For brevity, we call such a contract "detail-neutral." This term aims to highlight that certain details of the contracting situation do not matter for the functional form (7); it does not imply that the functional form is independent of all parameters. If $v(c) = c$ (the cost of effort is pecuniary) as assumed by HM, the contract is linear regardless of u, even though the cost function $g(a_t)$ may be nonlinear. The linear contracts of HM can thus be achieved in settings that do not require exponential utility, Gaussian noise or continuous time.

The origins of the contract's tractability and detail-neutrality can be seen in the heuristic proof. We first consider $T = 1$. Since η_1 is known, the expectations operator can be removed from (6). u then drops out to yield (9): u is irrelevant because it only affects the *magnitude* of the increment in utility that results from choosing the correct a_1 (i.e. the a_1 that solves (9)). This magnitude is irrelevant – the only important property is that it is always positive because u is monotonic. Regardless of the form that u takes, it is maximized by maximizing its argument, i.e. solving (9).

Even though all noise is known when the agent takes his action, it is not automatically irrelevant. First, since the agent does not know η_1 when he signs the contract, he is subject to risk and so the first-best is not achieved. Second, the noise realization has the potential to undo incentives. If there is a high η_1, r_1 and thus c will already be high, even if the agent exerts low effort. (A high reservation utility $\underline{u}$ has the same effect). If the agent exhibits diminishing marginal felicity (i.e. v is concave), he will have lower incentives to exert effort. The optimal contract must address this problem. It does so by being convex, via the v^{-1} transformation: if noise is high, it gives a greater number of dollars for exerting effort ($\frac{\partial c}{\partial r_1}$), to exactly offset the lower marginal felicity of each dollar ($v'(c)$). Therefore, the marginal felicity from effort remains $v'(c)\frac{\partial c}{\partial r_1} = g'(a_1^*)$, and incentives are preserved regardless of $\underline{u}$ or η. If the cost of effort is pecuniary ($v(c) = c$), $v^{-1}(c) = c$ and so no transformation is needed. Since both the costs and benefits of effort are in monetary terms, high past noise diminishes them equally. Thus, incentives are unchanged even with a linear contract.

We now move to the general case of $T > 1$. In all periods before the final one, the agent is now exposed to residual uncertainty, since he does not know future noise realizations when he chooses a_t. Much like the effect of a high current noise realization, if the agent expects future noise to be high, his incentives to exert effort will be reduced. In some models, this would require the agent to integrate over future noise realizations when choosing a_t. Here the unknown future noise outcomes do not matter, and this can be seen in the heuristic proof. Before $T+1$, η_{T+1} is unknown. However, (12) shows that the component of the contract that solves the $T+1$ problem ($g'\left(a^*_{T+1}\right) r_{T+1}$) is independent of that which solves the $t = 1, ..., T$ problems ($k\left(r_1, ..., r_T\right)$). Hence, the unknown η_{T+1} enters additively and does not affect the incentive constraints of the $t = 1, ..., T$ problems.[13] Our timing assumption thus allows us to solve the multiperiod problem via backward induction, reducing it to a succession of one-period problems, each of which can be solved separately. It has little effect on the economics of the situation since the agent continues to face uncertainty in all periods except for the final one, but instead is a technical device to allow us to collapse the problems.

Even though we can consider each problem separately, the periods remain interdependent. Much like the current noise realization, past outcomes may affect the current effort choice. The Mirrlees (1974) contract punishes the agent if final output is below a certain threshold. Therefore, if the agent can observe past outcomes, he will shirk if interim output is high. This complexity distinguishes our multiperiod model from a static multi-action model, where the agent chooses T actions simultaneously. As in HM, and unlike in a multi-action model, here the agent observes past outcomes when taking his current action, and can vary his action in response. HM assume exponential utility to remove such "wealth effects" and remove the intertemporal link between periods. We instead ensure that past outcomes do not distort incentives via the v^{-1} transformation described above, and so do not require exponential utility.

We achieve simple contracts in other settings than HM due to a different modeling framework. In HM, as in Grossman and Hart (1983), effort is modeled as the selection of a probability distribution over states of nature. Since effort only has a probabilistic effect on outcomes, the model already features uncertainty and so there is no need to introduce additional noise – noise dependence is not an issue. However, this formulation of effort requires exponential utility to remove wealth effects and achieve independence of the reservation wage.[14] By modeling effort as an increment to the signal (equation

[13] This can be most clearly seen in the definition of the new utility function (13), which "absorbs" the $T+1$ period problem.

[14] Specifically, the agent's objective function is $\sum_j u\left(c_j - g(p, \theta_j)\right) p_j$ where the summation is across states θ_j , and p_j is the probability of each state chosen by the agent. If the reservation wage is reduced by w, does the new incentive scheme simply subtract w from each c_j? The objective function would become

(1)), we achieve independence of u. This modeling choice requires the specification of a noise process, else the effort decision would become contractible. We then achieve tractability by our timing assumption, which allows us to solve the multiperiod problem by backward induction, and the v^{-1} transformation ensures incentives are preserved regardless of past noise or the reservation wage. In sum, the combination of the effort and timing specifications achieves tractability and independence of both utility u and noise η.

The Appendix proves that, even though the agent privately observes η_t, there is no need for him to communicate it to the principal. Since a_t^* is implemented for all η_t, there is a one-to-one correspondence between r_t and η_t on the equilibrium path. The principal can thus infer η_t from r_t, rendering messages redundant. The Appendix also shows that we can rule out randomized contracts. There are two effects of randomization. First, it leads to inefficient risk-sharing, for any concave u. Second, it alters the marginal cost of effort. If the utility function exhibits NIARA, this cost weakly increases with randomization. Thus, both effects of randomization are undesirable, and deterministic contracts are unambiguously optimal.[15] The proof makes use of the independence of noises and the log-concavity of $\eta_2, ..., \eta_T$. Note that while these assumptions, combined with the NIARA utility function, are sufficient to rule out randomized contracts, they may not be necessary. In future research, it would be interesting to explore whether randomized contracts can be ruled out in broader settings.[16]

In addition to allowing for stochastic contracts, the above analysis also allows for $a_t^* = \overline{a}$ for some t. When $a_t^* = \overline{a}$, the incentive compatibility constraint is an inequality. Therefore, the contract in (7) only provides a lower bound on the contract slope that implements a_t^*. A sharper-than-necessary contract has a similar effect to a stochastic contract, since it subjects the agent to additional risk. Again, the combination of NIARA and independent and log-concave noises is sufficient rule out such contracts. In sum, if the analysis allows for randomized contracts and $a_t^* = \overline{a}$, there are several incentive compatible contracts and the above three assumptions are sufficient to show that the contract in (7) is cheaper than stochastic contracts or contracts with a greater slope.

$\sum_j u\left(v(c_j) - w - c(p, \theta_j)\right) p_j$. Since p_j is outside the $u(\cdot)$ function, $u(\cdot)$ does not automatically drop out. Only if utility is exponential does the objective function simplify to $-u(-w) \sum_j u\left(v(c_j) - c(p, \theta_j)\right) p_j$, and so incentives are preserved by subtracting w from each payment c_j, i.e. the participation constraint can be met without affecting the incentive constraints. This property will not hold with non-exponential utility.

[15] This result builds on Arnott and Stiglitz (1988), who derived conditions under which randomization is suboptimal in a different setting of insurance.

[16] For instance, consider the case $T = 2$. We only require that $\widehat{u}(x)$ as defined in (40) exhibits NIARA. The concavity of η_2 is sufficient, but unnecessary for this. Separately, if NIARA is violated, the marginal cost of effort falls with randomization. However, this effect may be outweighed by the inefficient risk-sharing, so randomized contracts may still be dominated.

If the analysis is restricted to deterministic contracts and $a_t^* < \overline{a}\ \forall\ t$, the contract in (7) is the only incentive-compatible contract (for the signal values realized on the equilibrium path). We can thus drop the assumptions of NIARA utility and log-concave and independent noises. This result is stated in Proposition 1 below.

Proposition 1 *(Optimal deterministic contract, $a_t^* < \overline{a}\ \forall\ t$). Consider only deterministic contracts and $a_t^* < \overline{a}\ \forall\ t$. Drop the assumptions of NIARA utility, independent noises, and log-concave noises for $\eta_2, ..., \eta_T$; assume only that for all t, η_t has open interval support conditional on $\eta_1 ..., \eta_{t-1}$. Any incentive-compatible contract takes the form*

$$c = v^{-1}\left(\sum_{t=1}^{T} g'\left(a_t^*\right) r_t + K\right) \tag{15}$$

for some K, and for $\left(r_t\right)_{t\leq T}$ in the interior of the support of the realized values of the signal on the equilibrium path, $\left(a_t^* + \eta_t\right)_{t\leq T}$. Conversely, any contract that satisfies (15) for all r_t is incentive-compatible. The optimal contract has form (15), where K makes the agent's participation constraint bind.

Proof. See Appendix. ■

We close this section by considering two specific applications of Theorem 1 to executive compensation, to highlight the implications that stem from a tractable contract structure. While contract (7) can be implemented for any informative signal r, the firm's equity return is the natural choice of r for CEOs, since they are contracted to maximize shareholder value. When the cost of effort is pecuniary ($v\left(c\right) = c$), Theorem 1 implies that the CEO's dollar pay c is linear in the firm's return r. Hence, the relevant incentives measure is the dollar change in CEO pay for a given percentage change in firm value (i.e. "dollar-percent" incentives), as advocated by Hall and Liebman (1998).

Another common specification is $v(c) = \ln c$, in which case the CEO's utility function (2) now becomes, up to a monotonic (logarithmic) transformation:

$$\mathrm{E}\left[U\left(ce^{-g(a)}\right)\right] \geq \underline{U}, \tag{16}$$

where $u\left(x\right) \equiv U\left(e^{x}\right)$ and $\underline{U} \equiv \ln \underline{u}$ is the CEO's reservation utility. Utility is now multiplicative in effort and consumption salary; Edmans, Gabaix and Landier (2008) show that multiplicative preferences are necessary to generate empirically consistent predictions for the scaling of various measures of CEO incentives with firm size. Note that we retain the general utility function $U(\cdot)$.

Let r denote the firm's log return and $R = e^{r}$ denote its gross return. Applying Theorem 1 with $T = 1$ for simplicity, the optimal contract becomes

$$\ln c = g'(a^*)r + K. \tag{17}$$

The optimal contract is independent of the utility function U. It prescribes the percentage change in CEO pay for a percentage change in firm value, i.e. "percent-percent" incentives. Murphy (1999) advocated this elasticity measure over alternative incentive measures (such as "dollar-percent" incentives) on two empirical grounds: it is invariant to firm size, and firm returns have much greater explanatory power for percentage than dollar changes in pay. However, he notes that "elasticities have no corresponding agency-theoretic interpretation." The above analysis shows that elasticities are the theoretically justified measure under multiplicative preferences, regardless of the CEO's utility function. This result extends Edmans, Gabaix and Landier (2008) who advocated "percent-percent" incentives in a risk-neutral model.

2.2 Continuous Time

This section shows that the contract has the same detail-neutral functional form in continuous time, where actions and noise occur simultaneously. The consistency of the incentive scheme suggests that, if the underlying reality is continuous time, it is best mimicked in discrete time by modeling noise before the effort decision in each period.

At every instant t, the agent takes action a_t and the principal observes signal r_t, where

$$r_t = \int_0^t a_s ds + \eta_t, \tag{18}$$

$\eta_t = \int_0^t \sigma_s dZ_s + \int_0^t \mu_s ds$, Z_t is a standard Brownian motion, and $\sigma_t > 0$ and μ_t are deterministic. The agent's utility function is:

$$\mathrm{E}\left[u\left(v\left(c\right) - \int_0^T g\left(a_t\right) dt\right)\right]. \tag{19}$$

The principal observes the path of $(r_t)_{t\in[0,T]}$ and wishes to implement a deterministic action $(a_t^*)_{t\in[0,T]}$ at each instant. She solves Program 1 with utility function (19). The optimal contract is of the same detail-neutral form as Theorem 1.

Theorem 2 *(Optimal contract, continuous time). The following contract is optimal. The agent is paid*

$$c = v^{-1}\left(\int_0^T g'\left(a_t^*\right) dr_t + K\right), \tag{20}$$

where K is a constant that makes the participation constraint bind ($\mathrm{E}\left[u\left(\begin{array}{c}\int_0^T g'(a_t^*)\,dr_t + K\\ -\int_0^T g(a_t^*)\,dt\end{array}\right)\right] = \underline{u}$).

In particular, if the target action is time-independent ($a_t^* = a^* \; \forall \; t$), the contract

$$c = v^{-1}\left(g'(a^*)\, r_T + K\right) \tag{21}$$

is optimal.

Proof. See Appendix. ■

To highlight the link with the discrete time case, consider the model of Section 2.1 and define $r_T = \sum_{t=1}^T r_t = \sum_{t=1}^T a_t + \sum_{t=1}^T \eta_t$. Taking the continuous time limit of Theorem 1 gives Theorem 2.

2.3 Discussion: What is Necessary for Tractable, Detail-Neutral Contracts?

This section has extended the tractable contracts of HM to settings that do not require exponential utility, continuous time or Gaussian noise. However, the framework considered thus far has still imposed a number of restrictions, such as a risk-neutral principal, a linear signal, log-concave noises and a NIARA utility function. We now discuss the features that are essential for our contract structure, inessential features that we have already relaxed in extensions to the core model, and additional assumptions which may be relaxable in future research.

1. *Timing of noise.* We require that η_t is observed before a_t in each period. Without this assumption, the contract will depend on the utility function and the noise distribution. (The dependence on the noise distribution has been previously shown by Holmstrom (1979) and Grossman and Hart (1983), who assumed $u(x) = x$.) The Online Appendix shows that, even if a_t precedes η_t, contract (7) still implements $(a_t^*)_{t=1,..,T}$, although we can no longer show that it is optimal.

2. *Risk-neutral principal.* The proof of Theorem 1 extends the model to the case of a risk-averse principal. If the principal wishes to minimize $\mathrm{E}[w(c)]$ (where w is an increasing function) rather than $\mathrm{E}[c]$, then contract (7) is optimal if $u\left(v\left(w^{-1}(\cdot)\right) - \sum_t g(a_t^*)\right)$ is concave. This holds if, loosely speaking, the principal is not too risk-averse.

3. *NIARA utility, independent and log-concave noise.* Proposition 1 states that, if $a_t^* < \overline{a} \; \forall \; t$ and deterministic contracts are assumed, (7) is the only incentive-compatible contract. Therefore, these assumptions are not required.

4. *Unidimensional noise and action.* The Online Appendix shows that our model is readily extendable to settings where the action a and the noise η are multidimensional. A close analog to our result obtains.

5. *Linear signal.* The linearity of the signal, $r_t = a_t + \eta_t$, is not essential. Remark 1 in Section 3.1 later shows that with general signals $r_t = R(a_t, \eta_t)$, the optimal contract remains detail-neutral.

6. *Timing of consumption.* The current setup assumes that the agent only consumes at the end of period T. In ongoing work, we are developing the analog of Theorem 1 for repeated settings where the agent consumes in multiple periods. The key results remain robust to this extension.

7. *Renegotiation.* Since the target effort path is fixed, there is no scope for renegotiation when the agent observes the noise. In Section 3.1, the optimal action may depend on η. Since the optimal contract specifies an action for every realization of η, again there is no incentive to renegotiate.

3 The Optimal Effort Level

We have thus far assumed that the principal wishes to implement an exogenous path of effort levels (a_t^*). In Section 3.1 we allow the target effort level to depend on the noise. Section 3.2 shows that, in a broad class of situations, the principal will wish to implement the maximum productive effort level for all noise realizations (the "maximum effort principle").

3.1 Contingent Target Actions

Let $A_t(\eta_t)$ denote the "action function", which defines the target action for each noise realization. (Thus far, we have assumed $A_t(\eta_t) = a_t^*$, independent of η_t.) Since it is possible that different noises η_t could lead to the same observed signal $r_t = A_t(\eta_t) + \eta_t$, the analysis must consider revelation mechanisms; indeed, we find that the optimal contract now involves messages. If the agent announces noises $\widehat{\eta}_1, ..., \widehat{\eta}_T$, he is paid $c = C(\widehat{\eta}_1, ..., \widehat{\eta}_T)$ if the observed signals are $A_1(\widehat{\eta}_1) + \widehat{\eta}_1, ..., A_T(\widehat{\eta}_T) + \widehat{\eta}_T$, and a very low amount $\underline{c}$ otherwise.

As in the core model, we assume that $A_t(\eta_t) > \underline{a}\ \forall\ \eta_t$, else a flat contract would be optimal for some noise realizations. We make three additional technical assumptions: the action space $\mathcal{A}$ is open, $A_t(\eta_t)$ is bounded within any compact subinterval of η, and $A_t(\eta_t)$ is almost everywhere continuous. The final assumption still allows for a countable number of jumps in $A_t(\eta_t)$. Given the complexity and length of the proof that randomized contracts are inferior in Theorem 1, we now restrict the analysis to deterministic contracts and assume $A_t(\eta_t) < \overline{a}$. We conjecture that the same arguments in that proof continue to apply with a noise-dependent target action.

The optimal contract induces both the target effort level ($a_t = A_t(\eta_t)$) and truth-telling ($\widehat{\eta}_t = \eta_t$). It is given by the next Theorem:

Theorem 3 *(Optimal contract, noise-dependent action). The following contract is optimal. For each t, after noise η_t is realized, the agent communicates a value $\widehat{\eta}_t$ to the principal. If the subsequent signal is not $A_t(\widehat{\eta}_t) + \widehat{\eta}_t$ in each period, he is paid a very low amount $\underline{c}$. Otherwise he is paid $C(\widehat{\eta}_1, ..., \widehat{\eta}_T)$, where*

$$C(\eta_1, ..., \eta_T) = v^{-1}\left(\sum_{t=1}^{T} g(A_t(\eta_t)) + \sum_{t=1}^{T}\int_{\eta_*}^{\eta_t} g'(A_t(x))\,dx + K\right), \tag{22}$$

η_* is an arbitrary constant, and K is a constant that makes the participation constraint bind ($E\left[u\left(\sum_{t=1}^{T}\int_{\eta_*}^{\eta_t} g'(A_t(x))\,dx + K\right)\right] = \underline{u}$.)

Proof. (Heuristic). The Appendix presents a rigorous proof that does not assume differentiability of V and A. Here, we give a heuristic proof that conveys the essence of the result using first-order conditions. We set $T = 1$ and drop the time subscript.

Instead of reporting η, the agent could report $\widehat{\eta} \neq \eta$, in which case he receives $\underline{c}$ unless $r = A(\widehat{\eta}) + \widehat{\eta}$. Therefore, he must take action a such that $\eta + a = \widehat{\eta} + A(\widehat{\eta})$, i.e. $a = A(\widehat{\eta}) + \widehat{\eta} - \eta$. In this case, his utility is $V(\widehat{\eta}) - g(A(\widehat{\eta}) + \widehat{\eta} - \eta)$. The truth-telling constraint is thus:

$$\eta \in \arg\max_{\widehat{\eta}} V(\widehat{\eta}) - g(A(\widehat{\eta}) + \widehat{\eta} - \eta),$$

with first-order condition

$$V'(\eta) = g'(A(\eta))\,A'(\eta) + g'(A(\eta)).$$

Integrating over η gives the indirect felicity function

$$V(\eta) = g(A(\eta)) + \int_{\eta_*}^{\eta} g'(A(x))\,dx + K$$

for constants η_* and K. The associated pay is given by (22). ■

The functional form of the contract in Theorem 3 does not depend on $u(\cdot)$ nor on the distribution of η.[17] However, it is somewhat more complex than the contracts in Section 2,

[17] Even though (22) features an integral over the support of η, it does not involve the distribution of η.

as it involves calculating an integral. In the particular case where $A(\eta) = a^* \; \forall \; \eta$, Theorem 3 reduces to Theorem 1.

Remark 1 *(Extension of Theorem 3 to general signals). Suppose the signal is not $r_t = a_t + \eta_t$ but a general function $r_t = R(a_t, \eta_t)$, where R is differentiable and has positive derivatives in both arguments. The same analysis as in Theorem 3 derives the following contract as optimal:*

$$C(\eta_1, ..., \eta_T) = v^{-1}\left(\sum_{t=1}^{T} g(A_t(\eta)) + \int_{\eta_*}^{\eta_t} g'(A_t(x)) \frac{R_2(A_t(x), x)}{R_1(A_t(x), x)} dx + K\right), \tag{23}$$

where η_* is an arbitrary constant and K is a constant that makes the participation constraint bind.

The heuristic proof is as follows (setting $T = 1$ and dropping the time subscript). If η is observed and the agent reports $\widehat{\eta} \neq \eta$, he has to take action a such that $R(a, \eta) = R(A(\widehat{\eta}), \widehat{\eta})$. Taking the derivative at $\widehat{\eta} = \eta$ yields $R_1 \partial a / \partial \widehat{\eta} = R_1 A'(\eta) + R_2$. The agent solves $\max_{\widehat{\eta}} V(\widehat{\eta}) - g(a(\widehat{\eta}))$, with first-order condition $V'(\eta) - g'(A(\eta)) \partial a / \partial \widehat{\eta} = 0$. Substituting for $\partial a / \partial \widehat{\eta}$ from above and integrating over η yields (23).

3.2 Maximum Effort Principle for Large Firms

We now consider the optimal action function $A(\eta)$, specializing to $T = 1$ for simplicity. The principal chooses $A(\eta)$ to maximize

$$S \, \mathrm{E}\left[b\left(\min\left(A(\widetilde{\eta}), \overline{\overline{a}}\right), \widetilde{\eta}\right)\right] - \mathrm{E}\left[v^{-1}(V(\widetilde{\eta}))\right]. \tag{24}$$

The second term is the expected cost of compensation. It captures both the direct disutility from exerting effort $A(\eta)$ and the risk imposed by the incentive contract required to implement $A(\eta)$. The first term captures the productivity of effort, which is increasing in S, the baseline value of the output under the agent's control. For example, if the agent is a CEO, S is firm size; if he is a divisional manager, S is the size of his division. We will refer to S as firm size for brevity. Effort increases firm size to $S \, \mathrm{E}\left[b\left(\min\left(A(\widetilde{\eta}), \overline{\overline{a}}\right), \widetilde{\eta}\right)\right]$ where $b(\cdot)$ is the productivity function of effort and $\overline{\overline{a}} < \overline{a}$ is the maximum productive effort level. Defining $a = \min\left(A(\widetilde{\eta}), \overline{\overline{a}}\right)$, we assume that $b(a, \eta)$ is differentiable with respect to a, with $\inf_{a,\eta} \partial b(a, \eta) / \partial a > 0$. For example, if effort has a linear effect on the firm's log return, $b(a, \eta) = e^{a+\eta}$ and so effort affects firm value multiplicatively.

The min $\left(A(\widetilde{\eta}), \overline{\overline{a}}\right)$ function conveys the fact that, while the action space may be unbounded ($\overline{a}$ may be infinite), there is a limit to the number of productive activities the agent can undertake to benefit the principal. For example, if the agent is a CEO, there is a finite number of positive-NPV projects available; under the interpretation of a as rent extraction, $\overline{\overline{a}}$ reflects zero stealing. For all agents, there is a limit to the number of hours a day they can work while remaining productive. In addition to being economically realistic, this assumption is useful technically as it prevents the optimal action from being infinite. Actions $a > \overline{\overline{a}}$ do not benefit the principal, but improve the signal: one interpretation is manipulation (see Appendix C for further details). Clearly, the principal will never wish to implement $a > \overline{\overline{a}}$.

The next Theorem gives conditions under which maximum productive effort is optimal.[18]

Theorem 4 *(Optimality of maximum productive effort). Assume that* $\sup_{\left(\underline{a},\overline{\overline{a}}\right)} g''$ *and* $\sup_x \overline{F}(x)/f(x)$ *are finite, where* f *is the probability density function of* η*, and* $\overline{F}$ *is the complementary cumulative distribution (i.e.* $\overline{F}(x) = \Pr(\eta \geq x)$*). Define*

$$\Lambda = \left[\left(1 + \frac{u'(\alpha)}{u'(\beta)}\right)\left(\sup_{\left(\underline{a},\overline{\overline{a}}\right)} g''\right)\left(\sup \frac{\overline{F}}{f}\right) + g'\left(\overline{\overline{a}}\right)\right]\left(v^{-1}\right)'\left(\beta + g\left(\overline{\overline{a}}\right)\right), \tag{25}$$

where

$$\alpha \equiv u^{-1}(\underline{u}) - \left(\overline{\eta} - \underline{\eta}\right) g'\left(\overline{\overline{a}}\right) \text{ and } \beta \equiv u^{-1}(\underline{u}) + \left(\overline{\eta} - \underline{\eta}\right) g'\left(\overline{\overline{a}}\right).$$

When baseline firm size S *is above a threshold size* $S_* = \Lambda / \inf_{a,\eta} \frac{\partial b}{\partial a}(a,\eta)$*, implementing* $A(\eta) = \overline{\overline{a}}$ *is optimal for all* η*. Hence, allowing for noise-dependent actions, the optimal unrestricted contract is* $c = v^{-1}\left(g'\left(\overline{\overline{a}}\right) r + K\right)$*, where* K *is a constant that makes the agent's participation constraint bind.*

Proof. See Appendix. ■

The costs of effort are the disutility imposed on the CEO plus the risk imposed by incentives (summarized by Λ) and thus of similar order of magnitude to CEO pay. The benefits of effort are enhanced firm value and thus of similar order of magnitude to firm size. The productivity of effort also depends on the noise outcome, via the function $b(a,\eta)$. If the firm is sufficiently large ($S > S_*$), the benefits of effort outweigh the costs for all noise outcomes, and so dominate the trade-off. Therefore, maximum productive effort is optimal ($A(\eta) = \overline{\overline{a}}\ \forall\ \eta$.) A simple numerical example illustrates. Consider a firm with a \$20b market value and, to be conservative, assume that maximum CEO effort increases firm

[18] Theorem 4 states the assumption that $\sup_x \overline{F}(x)/f(x)$ is finite. A sufficient condition for this is to have f continuous, $f(x) > 0\ \forall\ x \in [\underline{\eta}, \overline{\eta})$, and f monotonic in a left neighborhood of $\overline{\eta}$. This condition is satisfied for many usual distributions.

value by only 1%. Then, maximum effort creates \$200m of value, which vastly outweighs the CEO's salary. Even if it is necessary to double the CEO's salary to compensate him for the costs of effort, the benefits of effort are much greater than the costs and so the firm wishes to implement maximum effort.

Combined with the results of Section 2, the optimal contract is detail-neutral in two dimensions – both the target effort level and the efficient implementation of this target. The irrelevance of risk is consistent with the empirical evidence surveyed by Prendergast (2002): a number of studies find that incentives are independent of risk, with the remainder equally divided between finding positive and negative correlations.

The comparative statics on the threshold firm size S_* are intuitive. First, S_* is increasing in noise dispersion, because the firm must be large enough for maximum effort to be optimal for all noise realizations. Indeed, a rise in $\overline{\eta} - \underline{\eta}$ increases β, lowers α, and raises $\sup \frac{\overline{F}}{f}$. (For example, if the noise is uniformly distributed, then $\sup \frac{\overline{F}}{f} = \overline{\eta} - \underline{\eta}$). Second, it is increasing in the agent's risk aversion and thus the risk imposed by incentives. For low noises, where the agent's utility is close to $\underline{u}$, $\frac{u'(\alpha)}{u'(\beta)} - 1$ is proportional to the agent's absolute risk aversion. Third, it is increasing in the disutility of effort, and thus the marginal cost of effort $g'\left(\overline{\overline{a}}\right)$ and the convexity of the cost function $\sup g''$. Fourth, it is decreasing in the marginal benefit of effort ($\inf_{a,\eta} \frac{\partial b}{\partial a}(a, \eta)$).

We conjecture that a "maximum effort principle" holds under more general conditions than those considered above. For instance, it likely continues to hold if the principal's objective function is $S\,\mathrm{E}\left[b\left(A\left(\widetilde{\eta}\right), \widetilde{\eta}\right)\right] - \mathrm{E}\left[v^{-1}\left(V\left(\widetilde{\eta}\right)\right)\right]$, and the action space is bounded above by $\overline{\overline{a}}$ – i.e. $\overline{a}$ (the maximum feasible effort level) equals $\overline{\overline{a}}$ (the maximum productive effort level). This slight variant is economically very similar, since the principal never wishes to implement $A(\eta) > \overline{\overline{a}}$ in our setting, but substantially more complicated mathematically, because the agent's action space now has boundaries and so the incentive constraints become inequalities. We leave the extension of this principle to future research.

3.3 Optimal Effort for Small Firms and Linear Cost of Effort

While Theorem 4 shows that $A(\eta) = \overline{\overline{a}}$ is optimal when $S > S_*$, we now show that $A(\eta)$ can be exactly derived even if $S \leq S_*$, if the cost function is linear – i.e. $g(a) = \theta a$, where $\theta > 0$.[19]

Proposition 2 *(Optimal contract with linear cost of effort). Let $g(a) = \theta a$, where $\theta > 0$.*

[19] Note 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 $g(a)$.

The following contract is optimal:

$$c = v^{-1}(\theta r + K), \tag{26}$$

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

$$A(\eta) \in \arg\max_{a \le \bar{\bar{a}}} Sb(a,\eta) - v^{-1}(\theta(a+\eta) + K). \tag{27}$$

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

Proof. From Theorem 3, 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:

$$\max_{A(\eta)} \mathrm{E}\left[Sb\left(\min\left(A(\eta), \bar{\bar{a}}\right), \eta\right)\right] - \mathrm{E}\left[v^{-1}(\theta(A(\eta)+\eta) + K)\right]$$

which is solved by pointwise maximization, as in (27). ∎

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 (27) is:

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

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

The main disadvantage is 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. Proposition 2 therefore assumes that indeterminacies are resolved by the agent following the principal's recommended action, $A(\eta)$.

4 Conclusion

This paper has identified and analyzed a class of situations in which the optimal contract is both tractable and detail-neutral. The contract can be solved in closed form and its slope is independent of the noise distribution and reservation utility; it is only determined by how the agent trades off the benefits of cash against the cost of effort. Furthermore, when the cost of effort can be expressed in financial terms, the optimal contract is linear, regardless of the utility function.

Holding the target effort level constant, detail-neutrality obtains in a multi-period discrete time model, where noise precedes effort in each period. The optimal contract is also the same in continuous time, where noise and actions occur simultaneously. Hence, if the underlying reality is continuous time, it is best mimicked in discrete time under our timing assumption. If the firm is sufficiently large, the target effort level is itself detail-neutral: the maximum effort level is optimal for a wide range of cost and effort functions and noise distributions. Since the benefits of effort are a function of total output, trade-off concerns disappear in a large firm, so maximum effort is efficient.

The model thus extends the tractable contracts of HM to settings that do not require exponential utility, continuous time or Gaussian noise. Moreover, it demonstrates which details of the contracting environment do and do not matter for the optimal incentive scheme. It can therefore rationalize why real-life contracts typically do not depend on as many specific details of the setting as existing literature might suggest – simply put, these details do not matter.

Our paper suggests several avenues for future research. The HM framework has proven valuable in many areas of applied contract theory owing to its tractability; however, some researchers have used the HM result in settings where the assumptions are not satisfied (see the critique of Hemmer (2004)). While we considered the specific application of executive compensation, other possibilities include bank regulation, team production, insurance or taxation. In particular, our contracts are valid in situations where time is discrete, utility cannot be modeled as exponential (e.g. in calibrated models where it is necessary to capture decreasing absolute risk aversion), or noise is not Gaussian (e.g. is bounded). In ongoing work, we are extending detail-neutral contracts to a dynamic setting where the agent consumes in each period, can privately save, and may smooth earnings intertemporally. In addition, while our model has relaxed a number of assumptions required for tractability, it continues to impose a number of restrictions. These are mostly technical rather than economic. For example, we feature a continuum of actions rather than a discrete set; our multiperiod model assumes independent noises with log-concave density functions; and our extension to noise-dependent target actions assumes an open action space and a maximum productive effort level. Some of these assumptions may not be valid in certain situations, limiting the applicability of our framework. Whether our setup can be further generalized is an open question for future research.

a	Effort (also referred to as "action")
$\overline{a}$	Maximum effort
$\overline{\overline{a}}$	Maximum productive effort
a^*	Target effort
b	Benefit function for effort, defined over a
c	Cash compensation
f	Density of the noise distribution
g	Cost of effort, defined over a
r	Signal (or "return"), typically $r = a + \eta$
u	Agent's utility function, defined over $v(c) - g(a)$
$\underline{u}$	Agent's reservation utility
v	Agent's felicity function, defined over c
η	Noise
A	Action function, defined over η
$\mathcal{C}$	Expected cost of contract
$\overline{F}$	Complementary cumulative distribution function for noise
M	Message sent by agent to the principal
S	Baseline size of output under agent's control
T	Number of periods
V	Felicity provided by contract

Table 1: Key Variables in the Model.

A Mathematical Preliminaries

This section derives some mathematical results that we use for the main proofs.

A.1 Dispersion of Random Variables

We repeatedly use the "dispersive order" for random variables to show that incentive compatibility 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 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 1 *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 2 *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:*

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

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 $\widehat{X} = -X$, $\widehat{Y} = -Y$, $\widehat{f}(x) = -f(-x)$, $h(x) = -\phi(-x)$. 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, $\mathrm{E}\left[\widehat{f}\left(\widehat{X}\right)\right] \geq \mathrm{E}\left[\widehat{f}\left(\widehat{Y}\right)\right]$. Thus, $\mathrm{E}\left[h\left(\widehat{f}\left(\widehat{X}\right)\right)\right] \geq \mathrm{E}\left[h\left(\widehat{f}\left(\widehat{Y}\right)\right)\right]$. Substituting $h\left(\widehat{f}\left(\widehat{X}\right)\right) = -\phi(f(X))$ yields $\mathrm{E}[-\phi(f(X))] \geq \mathrm{E}[-\phi(f(Y))]$. ■

Lemma 2 is intuitive: if $\mathrm{E}[f(X)] \geq \mathrm{E}[f(Y)]$, applying a concave function h should maintain the inequality. Conversely, if $\mathrm{E}[f(X)] \leq \mathrm{E}[f(Y)]$, applying a convex function ϕ should maintain the inequality. In addition, if $\mathrm{E}[X] = \mathrm{E}[Y]$, Lemma 2 implies that X second-order stochastically dominates Y. Hence, it is a stronger concept than second-order stochastic dominance.

Lemma 2 allows us to prove Lemma 3 below, which states that the NIARA property of a utility function is preserved by adding a log-concave random variable to its argument.

Lemma 3 *Let u denote a utility function with NIARA and Y a random variable with a log-concave distribution. Then, the utility function $\widehat{u}$ defined by $\widehat{u}(x) \equiv \mathrm{E}[u(x+Y)]$ 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

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

$\widehat{u}$ 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 $\mathrm{E}[u(b+Y+C_b)] = \mathrm{E}[u(b+Y+Z)]$. Hence, we can apply Lemma 2, which yields $\mathrm{E}[h(u(b+Y+C_b))] \geq \mathrm{E}[h(u(b+Y+Z))]$, i.e.

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

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.

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 maximum feasible effort level $\overline{a}$. 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. All the Lemmas in this subsection are proven in the Online Appendix.

Lemma 4 *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, $f(b) - f(a) \geq \int_a^b j(x)\,dx$.*

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

The following Lemma is the chain rule for subderivatives.

[20] For 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 $f(-1) > f(1)$.

Lemma 5 *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, $(h \circ f)'_{-}(x) = h'(f(x)) f'_{-}(x)$.*

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

Lemma 6 *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 $(f+h)'_{-}(x) = f'_{-}(x) + h'(x)$.*

B Detailed Proofs

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

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. This part of the proof is standard and can be skipped at a first reading. The second part proves the theorem considering only deterministic contracts and assuming that $a_t^* < \overline{a}\ \forall\ t$. This case requires weaker assumptions (see Proposition 1). The third part, which is significantly more complex, rules out randomized contracts and allows for the target effort to be the maximum $\overline{a}$. Both these extensions require the concepts of subderivatives and dispersion from Appendix A.

1). Redundancy of Messages

Let $\mathbf{r}$ denote the vector $(r_1, ..., r_T)$ and define $\boldsymbol{\eta}$ and $\mathbf{a}$ analogously. Define $\mathbf{g}(\mathbf{a}) = g(a_1) + ... + g(a_T)$. Let $\widetilde{V}_M(\mathbf{r}, \boldsymbol{\eta}) = v(\widetilde{c}(\mathbf{r}, \boldsymbol{\eta}))$ 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 $\widehat{\boldsymbol{\eta}} = \boldsymbol{\eta}$:

$$\forall \eta, \forall \widehat{\eta}, \forall \mathbf{a},\ \ \mathrm{E}\left[u\left(\widetilde{V}_M(\boldsymbol{\eta} + \mathbf{a}, \widehat{\boldsymbol{\eta}}) - \mathbf{g}(\mathbf{a})\right)\right] \leq \mathrm{E}\left[u\left(\widetilde{V}_M(\boldsymbol{\eta} + \mathbf{a}^*, \boldsymbol{\eta}) - \mathbf{g}(\mathbf{a}^*)\right)\right]. \quad (28)$$

The principal's problem is to minimize expected pay $\mathrm{E}\left[v^{-1}\left(\widetilde{V}_M\left(\widetilde{\boldsymbol{\eta}}+\mathbf{a}^*,\widetilde{\boldsymbol{\eta}}\right)\right)\right]$, subject to the IC constraint (28), and the agent's individual rationality (IR) constraint

$$\mathrm{E}\left[u\left(\widetilde{V}_M\left(\widetilde{\boldsymbol{\eta}}+\mathbf{a}^*,\widetilde{\boldsymbol{\eta}}\right)-\mathbf{g}\left(\mathbf{a}^*\right)\right)\right]\geq \underline{u}. \tag{29}$$

Since $\mathbf{r}=\mathbf{r}^*\equiv\mathbf{a}^*+\boldsymbol{\eta}$ on the equilibrium path, the message-dependent contract is equivalent to $\widetilde{V}_M\left(\mathbf{r},\mathbf{r}-\mathbf{a}^*\right)$. We consider replacing this with a new contract $\widetilde{V}\left(\mathbf{r}\right)$, 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:

$$\widetilde{V}\left(\mathbf{r}\right)=\widetilde{V}_M\left(\mathbf{r},\mathbf{r}-\mathbf{a}^*\right). \tag{30}$$

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

$$\forall\eta,\forall a,\ \ \mathrm{E}\left[u\left(\widetilde{V}\left(\mathbf{r}\right)-g\left(\mathbf{a}\right)\right)\right]\leq\mathrm{E}\left[u\left(\widetilde{V}\left(\mathbf{r}^*\right)-g\left(\mathbf{a}^*\right)\right)\right], \tag{31}$$

$$\mathrm{E}\left[u\left(\widetilde{V}\left(\mathbf{r}^*\right)-g\left(\mathbf{a}^*\right)\right)\right]\geq\underline{u}. \tag{32}$$

If the agent reports $\widehat{\boldsymbol{\eta}}\neq\boldsymbol{\eta}$, he must take action $\mathbf{a}$ such that $\boldsymbol{\eta}+\mathbf{a}=\widehat{\boldsymbol{\eta}}+\mathbf{a}^*$. Substituting $\widehat{\boldsymbol{\eta}}=\boldsymbol{\eta}+\mathbf{a}-\mathbf{a}^*$ into (28) and (29) indeed yields (31) and (32) 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 (30). Hence, the new contract $\widetilde{V}\left(\mathbf{r}\right)$ induces incentive compatibility and participation at the same cost as the initial contract $\widetilde{V}_M\left(\mathbf{r},\boldsymbol{\eta}\right)$ 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 $a_t^*<\overline{a}\ \forall\ t$

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:

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

Defining $r=\eta+a^*$ and $r'=\eta+a$, we have $a=a^*+r'-r$. The IC constraint can be rewritten:

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

Rewriting this inequality interchanging r and r' yields $g\left(a^*\right)-g\left(a^*+r-r'\right)\leq V\left(r'\right)-$

$V(r)$, and so:

$$g(a^*) - g(a^* + r' - r) \leq V(r) - V(r') \leq g(a^* + r - r') - g(a^*). \tag{33}$$

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

$$\frac{g(a^*) - g(a^* + r' - r)}{r - r'} \leq \frac{V(r) - V(r')}{r - r'} \leq \frac{g(a^* + r - r') - g(a^*)}{r - r'}. \tag{34}$$

Since a^* is in the interior of the action space $\mathcal{A}$ and the support of η is open, there exists r' in the neighborhood of r. Taking the limit $r' \uparrow r$, the first and third terms of (34) converge to $g'(a^*)$. Therefore, the left derivative $V'_{left}(r)$ exists, and equals $g'(a^*)$. Second, consider $r < r'$. Dividing (33) 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,

$$V'(r) = g'(a^*). \tag{35}$$

Since r has interval support[21], we can integrate to obtain, for some integration constant K:

$$V(r) = g'(a^*)\, r + K. \tag{36}$$

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 $a_t^* < \overline{a}$) does not require log-concavity of η_t, nor that u satisfies NIARA. This is because the contract (7) is the only incentive compatible contract. 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 (7).

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 $\widetilde{r}_1, ..., \widetilde{r}_T$ where $\widetilde{r}_2, ..., \widetilde{r}_T$ are log-concave, and a sequence of nonnegative numbers $g'(a_1^*), ..., g'(a_T^*)$. Consider the set of (potentially randomized) contracts $\widetilde{V}(r_1, ..., r_T)$ such*

[21] The 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.

that (i) $\mathrm{E}\left[u\left(\widetilde{V}\left(\widetilde{r}_1,...,\widetilde{r}_T\right)\right)\right] \geq \underline{u}$*; (ii)* $\forall\ t = 1...T$*,*

$$\frac{d}{d\varepsilon_-}\mathrm{E}\left[u\left(\widetilde{V}\left(\widetilde{r}_1,...,\widetilde{r}_t+\varepsilon,...,\widetilde{r}_T\right)\right) \mid \widetilde{r}_1,..,\widetilde{r}_t\right]_{|\varepsilon=0} \geq g'\left(a_t^*\right)\mathrm{E}\left[u'\left(\widetilde{V}\left(\widetilde{r}_1,...,\widetilde{r}_t,...,\widetilde{r}_T\right)\right) \mid \widetilde{r}_1,..,\widetilde{r}_t\right] \quad (37)$$

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

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

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

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

3a). Deterministic Contracts

The key difference from the proof in 2) is that we now must allow for $a_t^* = \overline{a}$.

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

(37) becomes $\frac{d}{d\varepsilon_-}u\left(V\left(r+\varepsilon\right)\right)_{|\varepsilon=0} \geq g'\left(a_1^*\right)u'\left(V\left(r\right)\right)$. Applying Lemma 5 to $h = u^{-1}$ yields:

$$V_-'\left(r\right) \geq g'\left(a^*\right). \quad (38)$$

It is intuitive that (38) 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 $g\left(a^*\right) = 0$ w.l.o.g.[22] If constraint (38) binds, the contract is $V^0\left(r\right) = g'\left(a^*\right)r + K$, where K satisfies $\mathrm{E}\left[u\left(g'\left(a^*\right)r+K\right)\right] = \underline{u}$. We wish to show that any other contract $V(r)$ that satisfies (38) is weaklier costlier.

By assumption (iii) in Statement Σ_1, V is nondecreasing. We can therefore apply Lemma 4 to equation (38), where condition (ii) of the Lemma is satisfied by $h\left(r\right) \equiv 0$. This implies that for $r \leq r'$, $V\left(r'\right) - V\left(r\right) \geq g'\left(a^*\right)\left(r'-r\right) = V^0\left(r'\right) - V^0\left(r\right)$. Thus, using Lemma 1, $V\left(\widetilde{r}\right)$ is more dispersed than $V^0\left(\widetilde{r}\right)$.

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

$$\mathrm{E}\left[u\left(V\left(\widetilde{r}\right)\right)\right] \geq \underline{u} = \mathrm{E}\left[u\left(V^0\left(\widetilde{r}\right)\right)\right]. \quad (39)$$

Applying Lemma 2 to the convex function $\phi \circ u^{-1}$ and inequality (39), we have:

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

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

i.e. $\mathrm{E}[\phi(V(\widetilde{r}))] \geq \mathrm{E}[\phi(V^0(\widetilde{r}))]$. 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 2.3, the extension to a risk-averse principal. Suppose that the principal wants to minimize $\mathrm{E}[w(c)]$, where w is an increasing and concave function, rather than $\mathrm{E}[c]$. 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.

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

$$\widehat{u}(x) = \mathrm{E}\left[u\left(x + g'\left(a^*_{T+1}\right)\widetilde{r}_{T+1}\right)\right]. \tag{40}$$

Since $\widetilde{r}_{T+1}$ is log-concave, $g'\left(a^*_{T+1}\right)\widetilde{r}_{T+1}$ is also log-concave. From Lemma 3, $\widehat{u}$ has the same NIARA property as u.

For each $\widetilde{r}_1, ..., \widetilde{r}_T$, we define $k(\widetilde{r}_1, ..., \widetilde{r}_T)$ as the solution to equation (41) below:

$$\widehat{u}(k(\widetilde{r}_1, ..., \widetilde{r}_T)) = \mathrm{E}_T[u(V(\widetilde{r}_1, ..., \widetilde{r}_{T+1}))]. \tag{41}$$

k represents the expected felicity from contract V based on all noise realizations up to and including 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 $\widehat{u}$ 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 $\widehat{u}$ and k at time T, thus allowing us to apply the Statement for these functions.

Taking expectations of (41) over $\widetilde{r}_1, ..., \widetilde{r}_T$ yields:

$$\mathrm{E}[\widehat{u}(k(\widetilde{r}_1, ..., \widetilde{r}_T))] = \mathrm{E}[u(V(\widetilde{r}_1, ..., \widetilde{r}_{T+1}))] \geq \underline{u}, \tag{42}$$

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 $\widehat{u}$ and contract k at time t. In addition, it is immediate that $\mathrm{E}[\widehat{u}(k(\widetilde{r}_1, ..., \widetilde{r}_T)) \mid \widetilde{r}_1, .., \widetilde{r}_t]$ is nondecreasing in $\widetilde{r}_t$. (Condition (iii)). We thus need to show that Condition (ii) is satisfied.

Since equation (37) holds for $t = T + 1$, we have

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

Applying Lemma 5 with function u yields:

$$\frac{dV}{dr_{T+1\,-}}\left(r_1, ..., r_{T+1}\right) \geq g'\left(a^*_{T+1}\right). \tag{43}$$

Hence, using Lemma 1 and Lemma 4, we see that conditional on $\widetilde{r}_1, ..., \widetilde{r}_T$, $V\left(\widetilde{r}_1, ..., \widetilde{r}_{T+1}\right)$ is more dispersed than $k\left(\widetilde{r}_1, ..., \widetilde{r}_T\right) + g'\left(a^*_{T+1}\right)\widetilde{r}_{T+1}$.

Using (40), we can rewrite equation (41) as

$$\mathrm{E}_T\left[u\left(k\left(\widetilde{r}_1, ..., \widetilde{r}_T\right) + g'\left(a^*_{T+1}\right)\widetilde{r}_{T+1}\right)\right] = \mathrm{E}_T\left[u\left(V\left(\widetilde{r}_1, ..., \widetilde{r}_{T+1}\right)\right)\right].$$

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 2 to yield:

$$\mathrm{E}_T\left[u' \circ u^{-1} \circ u\left(V\left(\widetilde{r}_1, ..., \widetilde{r}_{T+1}\right)\right)\right] \geq \mathrm{E}_T\left[u' \circ u^{-1} \circ u\left(k\left(\widetilde{r}_1, ..., \widetilde{r}_T\right) + g'\left(a^*_{T+1}\right)\widetilde{r}_{T+1}\right)\right], \text{ i.e.}$$
$$\mathrm{E}_T\left[u'\left(V\left(\widetilde{r}_1, ..., \widetilde{r}_{T+1}\right)\right)\right] \geq \mathrm{E}_T\left[\widehat{u}'\left(k\left(\widetilde{r}_1, ..., \widetilde{r}_T\right)\right)\right]. \tag{44}$$

Applying definition (41) to the left-hand side of Condition (ii) for $T + 1$ yields, with $t = 1...T$,

$$\frac{d}{d\varepsilon_-}\mathrm{E}_t\left[\widehat{u}\left(k\left(\widetilde{r}_1, ..., \widetilde{r}_t + \varepsilon, ..., \widetilde{r}_T\right)\right)\right]_{|\varepsilon=0} \geq g'\left(a^*_t\right)\mathrm{E}\left[u'\left(V\left(\widetilde{r}_1, ..., \widetilde{r}_t, ..., \widetilde{r}_{T+1}\right)\right) \mid \widetilde{r}_1, .., \widetilde{r}_t\right]$$

Taking expectations of equation (44) at time t and substituting into the right-hand side of the above equation yields:

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

Hence the IC constraint holds for contract $k\left(\widetilde{r}_1, ..., \widetilde{r}_T\right)$ 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 $k\left(r_1, ..., r_T\right)$, utility function $\widehat{u}$ and cost function $\widehat{\phi}$ defined by:

$$\widehat{\phi}(x) \equiv \mathrm{E}\left[\phi\left(x + g'\left(a_{T+1}\right)\widetilde{r}_{T+1}\right)\right]. \tag{45}$$

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

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

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

$$\mathcal{C}_k = \mathrm{E}\left[\widehat{\phi}\left(k\left(\widetilde{r}_1, ..., \widetilde{r}_T\right)\right)\right] \geq \mathcal{C}_{V^0} = \mathrm{E}\left[\phi\left(\sum_{t=1}^{T+1} g'(a_t^*) \widetilde{r}_t + K\right)\right]. \tag{46}$$

Using equation (45) yields:

$$\mathcal{C}_k = \mathrm{E}\left[\phi\left(k\left(\widetilde{r}_1, ..., \widetilde{r}_T\right) + g'(a_{T+1}) \widetilde{r}_{T+1}\right)\right] \geq \mathcal{C}_{V^0} = \mathrm{E}\left[\phi\left(\sum_{t=1}^{T+1} g'(a_t^*) \widetilde{r}_t + K\right)\right].$$

Finally, we compare the cost of contract $k(r_1, ..., r_T) + g'(a_{T+1}) \widetilde{r}_{T+1}$ to the cost of the original contract $V(r_1, ..., r_{T+1})$. Since equation (41) is satisfied, we can apply Lemma 2 to the convex function $\phi \circ u^{-1}$ and the random variable $\widetilde{r}_{T+1}$ to yield

$$\begin{aligned} \mathrm{E}_t\left[\phi\left(V\left(\widetilde{r}_1, ..., \widetilde{r}_{T+1}\right)\right)\right] &\geq \mathrm{E}_t\left[\phi\left(k\left(\widetilde{r}_1, ..., \widetilde{r}_T\right) + g'\left(a_{T+1}^*\right) \widetilde{r}_{T+1}\right)\right] \\ \mathrm{E}\left[\phi\left(V\left(\widetilde{r}_1, ..., \widetilde{r}_{T+1}\right)\right)\right] &\geq \mathrm{E}\left[\phi\left(k\left(\widetilde{r}_1, ..., \widetilde{r}_T\right) + g'\left(a_{T+1}^*\right) \widetilde{r}_{T+1}\right)\right] = \mathcal{C}_k \geq \mathcal{C}_{V^0}. \end{aligned}$$

where the final inequality comes from (46). 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 $\widetilde{V}(r_1, ..., r_T)$ and define the "certainty equivalent" contract $\overline{V}$ by:

$$u\left(\overline{V}(r_1, ..., r_T)\right) \equiv \mathrm{E}_T\left[u\left(\widetilde{V}(r_1, ..., r_T)\right)\right]. \tag{47}$$

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

From the above definition, we obtain

$$\mathrm{E}\left[u\left(\overline{V}\left(\widetilde{r}_1, ..., \widetilde{r}_T\right)\right)\right] = \mathrm{E}\left[u\left(\widetilde{V}\left(\widetilde{r}_1, ..., \widetilde{r}_T\right)\right)\right] \geq \underline{u},$$

i.e., $\overline{V}$ satisfies the participation constraint (29). Hence, Condition (i) holds. Also, it is clear that Condition (iii) holds for $\overline{V}$, given it holds for $\widetilde{V}$. We thus need to show that Condition (ii) is also satisfied. Applying Jensen's inequality to equation (47) and the function $u' \circ u^{-1}$ (which is convex since u exhibits NIARA) yields: $u'\left(\overline{V}(r_1, ..., r_T)\right) \leq$

$\mathrm{E}_T\left[u'\left(\widetilde{V}\left(r_1,...,r_T\right)\right)\right]$. We apply this to $r_t=\widetilde{r}_t$ for $t=1...T$ and take expectations to obtain

$$\mathrm{E}_t\left[u'\left(\widetilde{V}\left(\widetilde{r}_1,...,\widetilde{r}_T\right)\right)\right]\geq \mathrm{E}_t\left[u'\left(\overline{V}\left(\widetilde{r}_1,...,\widetilde{r}_T\right)\right)\right]. \tag{48}$$

Applying definition (47) to the left-hand side of (37) yields:

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

and using (48) yields:

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

Condition (ii) of Statement Σ_T therefore holds for $\overline{V}$. We can therefore apply Statement Σ_T to show that V^0 has a weakly lower cost than $\overline{V}$. We next show that the cost of $\overline{V}$ is weakly less than the cost of $\widetilde{V}$. Applying Jensen's inequality to (47) and the convex function $\phi\circ u^{-1}$ yields: $\phi\left(\overline{V}\left(r_1,...,r_T\right)\right)\leq \mathrm{E}\left[\phi\left(\widetilde{V}\left(r_1,...,r_T\right)\right)\right]$. We apply this to $r_t=\widetilde{r}_t$ for $t=1...T$ and take expectations over the distribution of $\widetilde{r}_t$ to obtain:

$$\phi\left(\overline{V}\left(\widetilde{r}_1,...,\widetilde{r}_T\right)\right)\leq \mathrm{E}\left[\phi\left(\widetilde{V}\left(\widetilde{r}_1,...,\widetilde{r}_T\right)\right)\right].$$

Hence $\overline{V}$ has a weakly lower cost than $\widetilde{V}$. Therefore, V^0 has a weakly lower cost than $\widetilde{V}$. 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 $\widetilde{r}_t\equiv a_t^*+\widetilde{\eta}_t$. We define

$$\overline{u}\left(x\right)\equiv u\left(x-\sum_{s=1}^{T}g\left(a_s^*\right)\right). \tag{49}$$

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

The IC constraint for time t is:

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

i.e.

$$0 \in \arg\max_{\varepsilon} \mathrm{E}_t\, u\left(\widetilde{V}\left(\widetilde{r}_1, ..., \widetilde{r}_t+\varepsilon, ..., \widetilde{r}_T\right) - g\left(a_t^*+\varepsilon\right) - \sum_{s=1...T, s\neq t} g\left(a_s^*\right)\right). \tag{50}$$

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 $\liminf_{y\uparrow 0}$, we obtain $\frac{d}{d\varepsilon_-} f'_-(\varepsilon)_{|\varepsilon=0} \geq 0$. Call $X(\varepsilon)$ the argument of u in equation (50). Applying this result to (50), we find: $\frac{d}{d\varepsilon_-} \mathrm{E}_t\, u\left(X(\varepsilon)\right)_{|\varepsilon=0} \geq 0$.

Using Lemma 5, we find $\mathrm{E}_t\left[u'(X(0))\left(\frac{d}{d\varepsilon_-} X(\varepsilon)_{|\varepsilon=0}\right)\right] \geq 0$. Using Lemma 6, $\frac{d}{d\varepsilon_-} X(\varepsilon)_{|\varepsilon=0} = \frac{d}{d\varepsilon_-} \widetilde{V}\left(\widetilde{r}_1, ..., \widetilde{r}_t+\varepsilon, ..., \widetilde{r}_T\right) - g'\left(a_t^*\right)$, hence we obtain:

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

Using again Lemma 5, this can be rewritten:

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

i.e., using the notation (49),

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

Therefore, Condition (ii) of Statement Σ_T holds.

Finally, we verify Condition (iii). Apply (50) to signal r_t and deviation $\varepsilon < 0$. We obtain:

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

so Condition (iii) holds for contract $\widetilde{V}$ and utility function $\overline{u}$.

We can now apply Statement Σ_T to contract $\widetilde{V}$ and function $\overline{u}$, to prove that any globally IC contract is weakly costlier than contract $V^0 = \sum_{t=1}^{T} g'\left(a_t^*\right) r_t + K$. Moreover, it is clear that V^0 satisfies the global IC conditions in equation (50). Thus, V^0 is the cheapest contract that satisfies the global IC constraint.

Proof of Proposition 1

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

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

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

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

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

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

This can be rewritten

$$a_T^* \in \arg\max_{a_T} \widehat{u}\left(K_T\left(a_1^*+\eta_1,...,a_T^*+\eta_T\right)-g\left(a_T\right)\right),$$

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

Using the same arguments as above for $T+1$, that implies that, for r_T in the interior of the support of $\widetilde{r}_T$ (given $(r_t)_{t\leq T-1}$) we can write:

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

for some function $K_{T-1}(r_1,...,r_T-1)$. 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 $(\widetilde{r}_t)_{t\leq T+1}$,

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

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:

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

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 Theorem 2

We shall use the following purely mathematical Lemma, proven in the Online Appendix.

Lemma 7 *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, $T \ge 0$, $X = \int_0^T \alpha_t dZ_t$, and $Y = \int_0^T \beta_t dZ_t$. Suppose that almost surely, $\forall t \in [0,T]$, $\alpha_t \le \beta_t$. Then X second-order stochastically dominates Y.*

Lemma 7 is intuitive: since $\beta_t \ge \alpha_t \ge 0$, 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 $(r_s)_{s\le t}$.

We proceed in two steps.

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

$$a_t^* \in \arg\max_{a_t} \theta_t a_t - g(a_t). \tag{51}$$

To prove this claim, consider another action policy (a_t), adapted to the filtration induced by $(Z_s)_{s\le t}$. 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 (51) 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 (51) is necessary for the contract to satisfy the IC condition.

We next show that condition (51) 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 $\theta_t = g'(a_t^*)$.

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

The case where $a_t^* \in (\underline{a}, \overline{a})\ \forall\ t$ 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 $a_t^* = \overline{a}$ 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

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

By reshifting $u(x) \to u\left(x - \int_0^T g(a_t^*)\,dt\right)$ 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 $\mathrm{E}[u(Y + K_Y)] \geq \underline{u}$, has a weakly greater expected cost than a contract $v = X + K_X$, with $\mathrm{E}[u(X + K_X)] = \underline{u}$. Lemma 7 implies that $\mathrm{E}[u(X+K_X)] \geq \mathrm{E}[u(Y+K_X)]$, and so

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

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 7 to function $-v^{-1}$ to yield:

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

where the second inequality follows from $K_X \leq K_Y$. Therefore, the expected cost of $v = X + K_X$ is weakly less that of $Y + K_Y$, and so contract $v = X + K_X$ is cost-minimizing. More explicitly, that is the contract (20) with $K = K_X + \int_0^T g'(a_t^*)\, a_t^* dt$.

Proof of Theorem 3

We prove the Theorem by induction.

Proof of Theorem 3 for $T = 1$. We remove time subscripts and let $V(\widehat{\eta}) = v(C(\widehat{\eta}))$ denote the felicity received by the agent if he announces $\widehat{\eta}$ and signal $A(\widehat{\eta}) + \widehat{\eta}$ is revealed.

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

$$V(\widehat{\eta}) - g(A(\widehat{\eta}) + \widehat{\eta} - \eta) \leq V(\eta) - g(A(\eta)). \tag{52}$$

Defining

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

the truth-telling constraint (52) can be rewritten,

$$g(A(\widehat{\eta})) - g(A(\widehat{\eta}) + \widehat{\eta} - \eta) \leq \psi(\eta) - \psi(\widehat{\eta}). \tag{53}$$

Rewriting this inequality interchanging η and $\widehat{\eta}$ and combining with the original inequality (53) yields:

$$\forall \eta, \forall \widehat{\eta} : g(A(\widehat{\eta})) - g(A(\widehat{\eta}) + \widehat{\eta} - \eta) \leq \psi(\eta) - \psi(\widehat{\eta}) \leq g(A(\eta) + \eta - \widehat{\eta}) - g(A(\eta)). \tag{54}$$

Consider a point η where A is continuous and take $\widehat{\eta} < \eta$. Dividing (54) by $\eta - \widehat{\eta} > 0$ and taking the limit $\widehat{\eta} \uparrow \eta$ yields $\psi'_{left}(\eta) = g'(A(\eta))$. Next, consider $\widehat{\eta} > \eta$. Dividing (54) by $\eta - \widehat{\eta} < 0$ and taking the limit $\widehat{\eta} \downarrow \eta$ yields $\psi'_{right}(\eta) = g'(A(\eta))$. Hence,

$$\psi'(\eta) = g'(A(\eta)), \tag{55}$$

at all points η where A is continuous.

Equation (55) 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 $\overline{a}_I = \sup \{A(\eta) + \eta - \widehat{\eta} \mid \eta, \widehat{\eta} \in I\}$, which is finite because A is assumed to be bounded in any compact subinterval of η. Then, equation (54) implies:

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

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 η, η_*, $\psi(\eta) = \psi(\eta_*) + \int_{\eta_*}^{\eta} \psi'(x)\, dx$. From (55), $\psi(\eta) = \psi(\eta_*) + \int_{\eta_*}^{\eta} g'(A(x))\, dx$, i.e.

$$V(\eta) = g(A(\eta)) + \int_{\eta_*}^{\eta} g'(A(x))\, dx + k \tag{56}$$

with $k = \psi(\eta_*)$. This concludes the proof for $T = 1$.

Proof that if Theorem 3 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 $\widehat{\eta}_{T+1}$, he must take action $a = A(\widehat{\eta}_{T+1}) + \widehat{\eta}_{T+1} - \eta_{T+1}$ so that the signal $a + \eta_{T+1}$ is consistent with declaring $\widehat{\eta}_{T+1}$. The IC constraint is therefore:

$$\eta_{T+1} \in \arg\max_{\widehat{\eta}_{T+1}} V(\eta_1, ..., \eta_T, \widehat{\eta}_{T+1}) - g(A(\widehat{\eta}_{T+1}) + \widehat{\eta}_{T+1} - \eta_{T+1}) - \sum_{t=1}^{T} g(a_t^*) . \tag{57}$$

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

$$V(\eta_1, ..., \eta_T, \widehat{\eta}_{T+1}) = W_{T+1}(\widehat{\eta}_{T+1}) + k(\eta_1, ..., \eta_T), \tag{58}$$

where $W_{T+1}(\widehat{\eta}_{T+1}) = g(A(\widehat{\eta}_{T+1})) + \int_{\eta_*}^{\widehat{\eta}_{T+1}} g'(A(x))\, dx$ and $k(\eta_1, ..., \eta_T)$ is the "constant" viewed from period $T+1$.

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

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

Defining

$$\widehat{u}(x) = \mathrm{E}\left[u\left(x + W_{T+1}\left(\widetilde{\eta}_{T+1}\right)\right)\right], \tag{59}$$

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

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

for some constant k_*. Combining this with (56), the only incentive compatible contract is:

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

Proof of Theorem 4

First, it is clear we can restrict ourselves to $A(\eta) \leq \overline{\overline{a}}$ for all η. If for some η, $A(\eta) > \overline{\overline{a}}$, the principal will be weakly better off by implementing $A(\eta) = \overline{\overline{a}}$ instead, since firm value $S\,\mathrm{E}\left[b\left(\min\left(A(\widetilde{\eta}), \overline{\overline{a}}\right), \widetilde{\eta}\right)\right]$ is unchanged, and the cost $\mathrm{E}\left[v^{-1}\left(V(\widetilde{\eta})\right)\right]$ will weakly decrease.

Let $\mathcal{C}(A)$ denote the expected cost of implementing $A(\eta)$, i.e. $\mathcal{C}(A) = \mathrm{E}\left[v^{-1}\left(C(\eta)\right)\right]$ where $C(\eta)$ is given by Theorem 3. The following Lemma states that the cost of effort is a Lipschitz-continuous function of the level of effort. Its proof is in the Online Appendix.

Lemma 8 *Suppose that g'' is bounded and that $\sup_x \overline{F}(x)/f(x) < \infty$. There is a constant Λ, given by equation (25) such that, for any two contracts that implement actions $A(\eta)$ and $B(\eta)$ in $(\underline{a}, \overline{\overline{a}}]$, the difference in the implementation costs satisfies: $|\mathcal{C}(A) - \mathcal{C}(B)| \leq \Lambda\,\mathrm{E}\,|A(\eta) - B(\eta)|$.*

By Lemma 8, we have $|\mathcal{C}^0 - \mathcal{C}| \leq \Lambda\,\mathrm{E}\left|\overline{\overline{a}} - A(\eta)\right|$. Next, let W^0 (respectively, W) denote the value of the principal's surplus (24) under the contract implementing $\overline{\overline{a}}$ (respectively, $A(\eta)$) and define $m = \inf_{a,\eta} \frac{\partial b}{\partial a}(a, \eta)$. The difference in total payoff to the firm is:

$$\begin{aligned} W^0 - W &= S\,\mathrm{E}\left[b\left(\overline{\overline{a}}, \eta\right)\right] - \mathcal{C}^0 - \left(S\,\mathrm{E}\left[b\left(A(\eta), \eta\right)\right] - \mathcal{C}\right) = S\,\mathrm{E}\left[b\left(\overline{\overline{a}}, \eta\right) - b\left(A(\eta), \eta\right)\right] - \left(\mathcal{C}^0 - \mathcal{C}\right) \\ &\geq Sm\,\mathrm{E}\left[\overline{\overline{a}} - A(\eta)\right] - \Lambda\,\mathrm{E}\left|\overline{\overline{a}} - A(\eta)\right| = (Sm - \Lambda)\,\mathrm{E}\left|\overline{\overline{a}} - A(\eta)\right|. \end{aligned}$$

Therefore, when $S > S_* \equiv \Lambda/m$, $W^0 - W > 0$ unless $\mathrm{E}\left[\left|\overline{\overline{a}} - A(\eta)\right|\right] = 0$. Hence, a contract that implements maximal effort for all noise realizations is optimal.

C A Microfoundation for the Principal's Objective

We offer a microfoundation for the principal's objective function (24). Suppose that the agent can take two actions, a "fundamental" action $a^F \in (\underline{a}, \overline{\overline{a}}]$ and a manipulative action $m \geq 0$. Firm value is a function of a^F only, i.e. the benefit function is $b\left(a^F, \eta\right)$. The signal is increasing in both actions: $r = a^F + m + \eta$. The agent's utility is $v(c) - \left[g^F(a) + G(m)\right]$, where g, G are increasing and convex, $G(0) = 0$, and $G'(0) \geq g'\left(\overline{\overline{a}}\right)$. The final assumption means that manipulation is costlier than fundamental effort.

We define $a = a^F + m$ and the cost function $g(a) = \min_{a^f, M} \left\{g^F(a) + G(m) \mid a^F + m = a\right\}$, so that $g(a) = g^F(a)$ for $a \in (\underline{a}, \overline{\overline{a}}]$ and $g(a) = g^F(a) + g(m - a)$ for $a \geq \overline{\overline{a}}$, which is increasing and convex. Then, firm value can be written $b\left(\min\left(a, \overline{\overline{a}}\right), \widetilde{\eta}\right)$, as in equation (24).

This framework is consistent with rational expectations. Suppose $b\left(a^F, \eta\right) = e^{a_F + \eta}$. After observing the signal r, the market forms its expectation P_1 of the firm value $b\left(a^F, \eta\right)$. The incentive contract described in Theorem 3 implements $a \leq \overline{\overline{a}}$, 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] Arnott, Richard and Joseph Stiglitz (1988) "Randomization with Asymmetric Information." *RAND Journal of Economics* 19, 344-362

[2] Baker, George (1992): "Incentive Contracts and Performance Measurement." *Journal of Political Economy* 100, 598-614

[3] Caplin, Andrew and Barry Nalebuff (1991): "Aggregation and Social Choice: A Mean Voter Theorem." *Econometrica* 59, 1-23

[4] DeMarzo, Peter and Michael Fishman (2007): "Optimal Long-Term Financial Contracting." *Review of Financial Studies* 20, 2079-2127

[5] DeMarzo, Peter and Yuliy Sannikov (2006): "Optimal Security Design and Dynamic Capital Structure in a Continuous-Time Agency Model." *Journal of Finance* 61, 2681-2724

[6] Dittmann, Ingolf, and Ernst Maug (2007): "Lower Salaries and No Options? On the Optimal Structure of Executive Pay." *Journal of Finance* 62, 303-343

[7] Dittmann, Ingolf, Ernst Maug and Oliver Spalt (2008): "Sticks or Carrots? Optimal CEO Compensation when Managers are Loss-Averse." Working Paper, Erasmus University Rotterdam

[8] Edmans, Alex, Xavier Gabaix and Augustin Landier (2008): "A Multiplicative Model of Optimal CEO Incentives in Market Equilibrium." *Review of Financial Studies*, forthcoming

[9] Grossman, Sanford and Oliver Hart (1983): "An Analysis of the Principal-Agent Problem." *Econometrica* 51, 7-45

[10] Hall, Brian and Jeffrey Liebman (1998): "Are CEOs Really Paid Like Bureaucrats?" *Quarterly Journal of Economics*, 113, 653-691

[11] Hall, Brian and Kevin Murphy (2002): "Stock Options for Undiversified Executives." *Journal of Accounting and Economics* 33, 3-42

[12] Harris, Milton and Artur Raviv (1979): "Optimal Incentive Contracts With Imperfect Information." *Journal of Economic Theory* 20, 231-259

[13] Hellwig, Martin and Klaus Schmidt (2002): "Discrete-Time Approximations of the Holmstrom-Milgrom Brownian-Motion Model of Intertemporal Incentive Provision." *Econometrica* 70, 2225-2264

[14] Hemmer, Thomas (2004): "Lessons Lost in Linearity: A Critical Assessment of the General Usefulness of LEN Models in Compensation Research." *Journal of Management Accounting Research* 16, 149-162

[15] Holmstrom, Bengt and Paul Milgrom (1987): "Aggregation and Linearity in the Provision of Intertemporal Incentives." *Econometrica* 55, 308-328

[16] Jewitt, Ian (1988): "Justifying the First-Order Approach to Principal-Agent Problems." *Econometrica* 56, 1177-1190

[17] Karatzas, Ioannis and Steven E. Shreve (1991): *Brownian Motion and Stochastic Calculus*, 2nd edition, Springer Verlag

[18] Krishna, Vijay and Eliot Maenner (2001): "Convex Potentials with an Application to Mechanism Design." *Econometrica* 69,1113-1119

[19] Lacker, Jeffrey and John Weinberg (1989): "Optimal Contracts under Costly State Falsification." *Journal of Political Economy* 97, 1345-1363

[20] Laffont, Jean-Jacques and Jean Tirole (1986): "Using Cost Observation to Regulate Firms." *Journal of Political Economy* 94, 614-641

[21] Landsberger, Michael and Isaac Meilijson (1994): "The Generating Process and an Extension of Jewitt's Location Independent Risk Concept." *Management Science* 40, 662-669

[22] Mirrlees, James (1974): "Notes on Welfare Economics, Information and Uncertainty" in Michael Balch, Daniel McFadden, and Shih-Yen Wu, eds., *Essays on Economic Behavior under Uncertainty*, North-Holland, Amsterdam.

[23] Mueller, Holger (2000): "Asymptotic Efficiency in Dynamic Principal-Agent Problems." *Journal of Economic Theory* 91, 292-301

[24] Murphy, Kevin (1999): "Executive Compensation" in Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics*, Vol. 3b. New York and Oxford: Elsevier/North-Holland, 2485-2563

[25] Ou-Yang, Hui (2003): "Optimal Contracts in a Continuous-Time Delegated Portfolio Management Problem." *Review of Financial Studies* 16, 173-208

[26] Prendergast, Canice (2002): "The Tenuous Trade-Off between Risk and Incentives." *Journal of Political Economy*, 110, 1071-102

[27] Rogerson, William (1985): "The First Order Approach to Principal-Agent Problems." *Econometrica* 53, 1357-1368

[28] Rudin, Walter (1987): *Real and Complex Analysis*, 3rd edition, McGraw-Hill

[29] Sannikov, Yuliy (2008): "A Continuous-Time Version of the Principal-Agent Problem." *Review of Economic Studies*, 75, 957-984.

[30] Schaettler, Heinz and Jaeyoung Sung (1993): "The First-Order Approach to the Continuous-Time Principal-Agent Problem With Exponential Utility." *Journal of Economic Theory* 61, 331-371

[31] Shaked, Moshe and George Shanthikumar (2007): *Stochastic Orders*, Springer Verlag

[32] Sung, Jaeyoung (1995): "Linearity with Project Selection and Controllable Diffusion Rate in Continuous-Time Principal-Agent Problems." *RAND Journal of Economics* 26, 720-743

[33] Wang, Susheng (2007): "The Limit Contract in the Standard Agency Model." Working Paper, Hong Kong University of Science and Technology

[34] Wilson, Robert (1987): "Game Theoretic Approaches to Trading Processes" in Truman Bewley, ed., *Advances in Economic Theory: Fifth World Congress*, Cambridge University Press, Cambridge

Online Appendix for "Tractability and Detail-Neutrality in Incentive Contracting"

Alex Edmans and Xavier Gabaix
November 12, 2008

D Incentive Compatibility of Contract when Timing is Reversed

In the core model, noise η_t precedes the action a_t in each period. This section shows that, if the timing is reversed, the optimal contract in Theorem 1 still induces the target path of actions, although we can no longer prove that it is incentive compatible. For brevity, we consider $T = 1$.

The agent chooses

$$a^* \in \arg\max_a \mathrm{E}\left[u\left(v\left(c\left(a+\eta\right)\right) - g\left(a\right)\right)\right],$$

where η is now unknown. With the proposed contract $v\left(c\left(r\right)\right) = g'\left(a^*\right) r + K$, so the maximization problem is:

$$a^* \in \arg\max_a \mathrm{E}\left[u\left(g'\left(a^*\right) a - g\left(a\right) + g'\left(a^*\right) \eta\right)\right]$$

This is maximized pointwise by maximizing $g'\left(a^*\right) a - g\left(a\right)$ over a, i.e. for $a = a^*$.

However, we can no longer prove that the contract in Theorem 1 is optimal. In general, results from Holmstrom (1979) indicate that it is not optimal with that "reversed" timing.

E 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:

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

where $\boldsymbol{\eta}, \mathbf{r} \in \mathbb{R}^S$, and $\mathbf{b}{:}\mathcal{A} \in \mathbb{R}^I \to \mathbb{R}^S$. The signal and action can be of different dimensions. In the core model, $S = I = 1$ and $\mathbf{b}(a) = a$. 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 3 *(Optimal contract, discrete time, multidimensional signal and action). Define the $I \times S$ matrix $L = \mathbf{b}'(\mathbf{a}^*)^\top$ i.e. explicitly $L_{ij} = \frac{\partial \mathbf{b}_j}{\partial a_i}(a_1^*, ..., a_I^*)$, and assume that there is a vector $\theta \in \mathbb{R}^S$ such that*

$$L\theta = g'(\mathbf{a}^*), \tag{60}$$

i.e., explicitly:

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

The optimal contract is given by:

$$c(\mathbf{r}) = v^{-1}(\theta\mathbf{r} + K(\mathbf{r})), \tag{61}$$

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

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

$$\forall \mathbf{r}, LK'(\mathbf{r}) = 0 \tag{62}$$

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

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:

$$V'(\mathbf{b}(\mathbf{a}^*) + \boldsymbol{\eta})\,\mathbf{b}'(\mathbf{a}^*) - g'(\mathbf{a}^*) = 0, \tag{63}$$

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 equation (63) 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. Previously, K was 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:

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

In this case, $L = (1\ 1)$. Therefore, with $\theta = (\theta_1, \theta_2) \in \mathbb{R}^2$, (60) becomes: $\theta_1 + \theta_2 = g'(a^*)$. For example, we can take $\theta_1 = \theta_2 = g'(a^*)/2$. Next, (62) 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):

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

In this case, $L = (1\ 0)$. Therefore, with $\theta = (\theta_1, \theta_2) \in \mathbb{R}^2$, (60) becomes: $\theta_1 = g'(a^*)$. Next, (62) 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:

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

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.

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

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, $r_n(x) \geq \min\left(0, -\sup_{[a,b]} h'(x)\right)\ \forall\ x$. Hence we can apply Fatou's lemma, which shows:

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

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 $\liminf_{n\to\infty} A_n \geq 0$.

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 $\liminf_{n\to\infty} B_n \geq 0$.

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

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

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

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

If $\lim_{n\to\infty} f(y_n) > f(x)$, 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 $\lim_{n\to\infty} f(\hat{y}_n)$ exists. If $\lim_{n\to\infty} f(\hat{y}_n) = f(x)$, Then,

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

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

If $\lim_{n\to\infty} f(\hat{y}_n) > f(x)$, 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 $\mathrm{E}[h(X)] \geq \mathrm{E}[h(Y)]$ for any concave function h. Define $I(\delta) = \mathrm{E}[h(X+\delta(Y-X))]$ for $\delta \in [0,1]$, so that

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

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:

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

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:

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

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.

Proof of Lemma 8

We commence the proof with a Lemma.

Lemma 9 *Consider a differentiable function f, two continuously differentiable random variables X and Y (not necessarily independent), two constants a and b such that $\mathrm{E}[f(X+a)] = \mathrm{E}[f(Y+b)]$, and an interval I such that (i) $f'(x) > 0 \;\forall\; x \in I$ and (ii) almost surely, $X+a$ and $Y+b$ are in I. Then,*

$$|a-b| \leq \frac{\sup_I f'}{\inf_I f'} \mathrm{E}[|X-Y|]. \tag{64}$$

The Lemma implies that when X and Y are "close", then a and b are also close.

Proof. By redefining if necessary $Y' = Y + b$, $X' = X + b$, it is sufficient to consider the case $b = 0$. Define $H = X - Y + a$. From the Intermediate Value Theorem, for any Y, H, there is a value $\theta(Y,H)$ such that $f(Y+H) - f(Y) = f'(Y + \theta(Y,H)\,H)\,H$. In addition, $Y + \theta(Y,H)\,H \in I$ almost surely. Hence,

$$\begin{aligned}0 &= \mathrm{E}[f(Y)] - \mathrm{E}[f(X+a)] = \mathrm{E}[f(Y)] - \mathrm{E}[f(Y+H)] = \mathrm{E}[f'(Y+\theta(Y,H)\,H)(X-Y+a)] \\ &= \mathrm{E}[f'(Y+\theta(Y,H)\,H)(X-Y)] + a\,\mathrm{E}[f'(Y+\theta(Y,H)\,H)].\end{aligned}$$

Thus,

$$|a| = \frac{|\mathrm{E}[f'(Y+\theta(Y,H)\,H)(X-Y)]|}{\mathrm{E}[f'(Y+\theta(Y,H)\,H)]} \leq \frac{(\sup_I f')\,\mathrm{E}[|X-Y|]}{\inf_I f'}.$$

■

We now turn to the main proof. Consider contract A that implements action $A(\eta)$, and contract B that implements $B(\eta)$. Define

$$X = \int_{\underline{\eta}}^{\eta} g'(A(x))\,dx, \qquad Y = \int_{\underline{\eta}}^{\eta} g'(B(x))\,dx,$$

and k, k' such that

$$\underline{u} = \mathrm{E}[u(X+k)] = \mathrm{E}[u(Y+k')]. \tag{65}$$

From Proposition 3, the felicity of contract A is $X + k + g(A(\eta))$, and the felicity of contract B is $Y + k' + g(B(\eta))$.

We prove the Lemma by demonstrating a sequence of three inequalities.

1). Inequality regarding $|k - k'|$**.** Since $0 \leq X \leq (\overline{\eta} - \underline{\eta})\, g'(\overline{\overline{a}})$, we have

$$u(k) \leq \underline{u} = \mathrm{E}[u(X+k)] \leq u\left(k + (\overline{\eta} - \underline{\eta})\, g'(\overline{\overline{a}})\right)$$
$$\Rightarrow u^{-1}(\underline{u}) - (\overline{\eta} - \underline{\eta})\, g'(\overline{\overline{a}}) \leq k \leq u^{-1}(\underline{u}).$$

We therefore have $\alpha \leq k + X \leq \beta$, where

$$\alpha \equiv u^{-1}(\underline{u}) - (\overline{\eta} - \underline{\eta})\, g'(\overline{\overline{a}}) \text{ and } \beta \equiv u^{-1}(\underline{u}) + (\overline{\eta} - \underline{\eta})\, g'(\overline{\overline{a}}). \tag{66}$$

By the same reasoning, $\alpha \leq k' + Y \leq \beta$.

Applying Lemma 9 to equation (65), function u and interval $[\alpha, \beta]$, we obtain: $|k - k'| \leq \frac{\sup_{[\alpha,\beta]} u'}{\inf_{[\alpha,\beta]} u'} \mathrm{E}|X - Y|$. Since u is concave, this yields the inequality:

$$|k - k'| \leq \frac{u'(\alpha)}{u'(\beta)} \mathrm{E}|X - Y|. \tag{67}$$

2). Inequality regarding $\mathrm{E}|X - Y|$. We have:

$$\begin{aligned}
\mathrm{E}|X - Y| &= \mathrm{E}\left|\int_{\underline{\eta}}^{\widetilde{\eta}} \left(g'(A(x)) - g'(B(x))\right) dx\right| \\
&\leq (\sup g'')\, \mathrm{E}\left[\int_{\underline{\eta}}^{\widetilde{\eta}} |A(x) - B(x)|\, dx\right] = (\sup g'')\, \mathrm{E}\left[\int_{\underline{\eta}}^{\overline{\eta}} |A(x) - B(x)|\, 1_{x \leq \widetilde{\eta}} dx\right] \\
&= (\sup g'') \int_{\underline{\eta}}^{\overline{\eta}} |A(x) - B(x)|\, \mathrm{E}[1_{x \leq \widetilde{\eta}}]\, dx \\
&= (\sup g'') \int_{\underline{\eta}}^{\overline{\eta}} |A(x) - B(x)|\, \overline{F}(x)\, dx, \text{ defining } \overline{F}(x) = P(\eta \geq x) \\
&= (\sup g'') \int_{\underline{\eta}}^{\overline{\eta}} |A(x) - B(x)| \frac{\overline{F}(x)}{f(x)} f(x)\, dx \\
&\leq (\sup g'') \left(\sup \frac{\overline{F}}{f}\right) \int_{\underline{\eta}}^{\overline{\eta}} |A(x) - B(x)|\, f(x)\, dx = (\sup g'') \left(\sup \frac{\overline{F}}{f}\right) \mathrm{E}|A(\widetilde{\eta}) - B(\widetilde{\eta})|,
\end{aligned}$$

yielding the inequality

$$\mathrm{E}|X - Y| \leq (\sup g'') \left(\sup \frac{\overline{F}}{f}\right) \mathrm{E}|A(\widetilde{\eta}) - B(\widetilde{\eta})|. \tag{68}$$

3). Inequality regarding the difference in costs. We can now compare the costs of the two contracts, which we denote $\mathcal{C}_A$ and $\mathcal{C}_B$. We find:

$$
\begin{aligned}
|\mathcal{C}_A - \mathcal{C}_B| &= \left|\mathrm{E}\left[v^{-1}\left(X + k + g\left(A\left(\eta\right)\right)\right) - v^{-1}\left(Y + k + g\left(B\left(\eta\right)\right)\right)\right]\right| \\
&\leq \Big(\sup_{[\alpha + \inf g, \beta + \sup g]} \left(v^{-1}\right)'\Big) \cdot \mathrm{E}\left[\left|X + k + g\left(A\left(\eta\right)\right) - \left(Y + k' + g\left(B\left(\eta\right)\right)\right)\right|\right] \\
&\leq D\left(\mathrm{E}\left|X - Y\right| + \left|k - k'\right| + \mathrm{E}\left[g\left(A\left(\eta\right) - g\left(B\left(\eta\right)\right)\right)\right]\right), \text{ defining } D = \left(v^{-1}\right)'\left(\beta + g\left(\overline{\overline{a}}\right)\right) \\
&\leq D\left(1 + \frac{u'\left(\alpha\right)}{u'\left(\beta\right)}\right)\mathrm{E}\left|X - Y\right| + Dg'\left(\overline{\overline{a}}\right)\mathrm{E}\left|A\left(\eta\right) - B\left(\eta\right)\right|, \text{ by equation (67)}
\end{aligned}
$$

Define

$$
\Lambda = \left[\left(1 + \frac{u'\left(\alpha\right)}{u'\left(\beta\right)}\right)\left(\sup g''\right)\left(\sup \frac{\overline{F}}{f}\right) + g'\left(\overline{\overline{a}}\right)\right]\left(v^{-1}\right)'\left(\beta + g\left(\overline{\overline{a}}\right)\right), \qquad (69)
$$

where α, β are given in equation (66), and $\overline{F}\left(x\right) = P\left(\eta \geq x\right)$. Using equation (68) yields:

$$
\left|\mathcal{C}_A - \mathcal{C}_B\right| \leq \Lambda\,\mathrm{E}\left|A\left(\eta\right) - B\left(\eta\right)\right|,
$$

as required.

References

[1] Di Nunno, Guilia, Berndt Oksendal and Frank Proske (2008): *Malliavin Calculus for Lévy Processes with Applications to Finance*, Springer Verlag

[2] Holmstrom, Bengt (1979): "Moral Hazard and Observability." *Bell Journal of Economics* 10, 74-91

[3] Holmstrom, Bengt (1982): "Moral Hazard in Teams." *Bell Journal of Economics* 13, 324-340