Detail-Independent Contracting, with an Application to CEO Incentives*

Alex Edmans — Wharton

Xavier Gabaix — NYU Stern and NBER

September 17, 2008

Abstract

This paper identifies a broad class of situations in which the contract is both attainable in closed form and "detail-independent". Its slope is independent of firm risk; moreover, when the cost of effort is pecuniary, the contract is linear regardless of the agent's utility function. We thus extend the tractable contracts of Holmstrom and Milgrom (1987) to settings that do not require exponential utility, Gaussian noise or continuous time. In particular, the optimal continuous time contract is also efficient in a discrete model. Our results are consistent with the simplicity of real-life contracts, and suggest that incentive schemes need not depend on complex details of the particular setting (e.g. agent's risk aversion), which are difficult for the principal to observe. For CEOs, the contract can be implemented with securities in a simple manner. The CEO is given an "Incentive Account": a portfolio that is continuously rebalanced so that the fraction invested in the firm's stock remains above a certain threshold. This threshold is also independent of noise and utility.

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 Heski Bar-Isaac, Ingolf Dittmann, Ernst Maug, Canice Prendergast, Rob Tumarkin, Yuliy Sannikov, seminar participants at Chicago GSB, Northwestern, NYU, and Wisconsin. 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

This paper identifies and analyzes a class of contracting situations in which the optimal incentive scheme is both attainable in closed form and "detail-independent." The contract slope is independent of the noise distribution and the agent's reservation utility. Moreover, if the cost of effort is in monetary terms (e.g. it represents an opportunity cost of working elsewhere), the contract is linear – regardless of the agent's utility function. The model thus extends the tractable contracts of Holmstrom and Milgrom (1987)[1] to settings that do not require exponential utility, Gaussian noise or continuous time. Our structure may be useful for future contracting models, as it allows tractability to be achieved in a wide range of settings. Moreover, it suggests why real-life contracts are typically simple and do not depend on specific details of the contracting situation.

We commence by assuming a deterministic (but possibly time-varying) path of target effort levels, and analyze the cheapest incentive scheme that implements this path. We 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. The contract is both tractable and detail-independent: the contract slope depends only on how the agent trades off the benefits of cash against the cost of providing effort, and not on any other factors.[2] The irrelevance of the noise distribution occurs even though each action, except the final one, is followed by noise and so the agent faces uncertainty when deciding his effort level. Moreover, if the timing is reversed, so that actions precede noise in each period, the contract still implements the target action, although it may no longer be optimal. We then use recent advances in continuous time contracting (Sannikov (2008)) to show that the optimal contract is the same in a continuous-time model where noise and effort occur simultaneously.

We next endogenize the target effort path and, in particular, allow it to depend on the noise outcomes. The optimal contract is still attainable in closed form. 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). We show that, if the output under the agent's

[1] Throughout this paper, unless otherwise stated, "Holmstrom and Milgrom (1987)" refers to the continuous-time, closed form linear contract derived in Section 4 of their paper.

[2] For brevity, we call such a contract "detail-independent." This term emphasizes that certain details of the contracting situation do not matter (in contrast to earlier theories). However, it does not mean that all parameters are irrelevant.

control is sufficiently large (e.g. the agent is a CEO who affects total firm value), these trade-off considerations are of second-order importance: the benefits of effort swamp the costs. Thus, maximum effort is optimal in each period for a wide range of utility functions and noise outcomes. By contrast, if output is small, maximum effort may not be optimal for some noise realizations. We show that the optimal effort level can still be solved for if the cost function is affine.

In sum, for a given target effort level, the optimal implementation is detail-independent. Moreover, if output is sufficiently large, the optimal action itself does not depend on model parameters, and so the overall contract is also detail-independent. 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.

The "maximum effort principle"[3], when applicable, significantly increases tractability, since it removes the need to solve the trade-off required to derive an interior optimum. Indeed, jointly deriving the optimal effort level and the efficient contract that implements it can be extremely complex. Thus, papers that analyze the second (implementation) problem typically assume a fixed target effort level (e.g. Grossman and 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.

We demonstrate how the model can be applied to the design of CEO incentives. Since CEOs affect overall firm value, the contract can be implemented using firm securities. If CEO preferences are multiplicative in cash and effort (which Edmans, Gabaix and Landier (2008) show to be necessary for empirically consistent predictions), the implementation with securities takes a simple form. At the start, the CEO is given an "Incentive Account": a portfolio for which a given fraction is invested in the firm's stock and the remainder in cash. As time evolves, this portfolio is continuously rebalanced, so that the fraction in the firm's stock remains above a given threshold, and the CEO is paid the final portfolio value in the last period. This threshold fraction depends only on the unit cost of effort and not on noise or utility. The Incentive Account resembles restricted stock and long-vesting options frequently granted in practice. Existing justifications of long-term compensation are typically based on myopia or manipula-

[3] 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.

tion concerns, in which case the optimal vesting point is time-dependent. By contrast, in this paper, restricted compensation is desirable to maintain effort incentives in a dynamic model. In this case, the optimal vesting point is state-dependent: the CEO should only be allowed to cash out if the fraction is sufficiently higher than the threshold, which in turn requires the stock price to have increased. On the other hand, if the stock price falls, the requirement to maintain a minimum incentive level may justify the repricing of out-of-the-money options.

In addition to executive compensation, our framework may also be applicable to theoretical contracting models in general, as it shows that tractability can be achieved in quite broad settings. For example, certain models may require decreasing relative risk aversion to calibrate to the data and/or discrete time for clarity. In addition, the results also have implications for empirical researchers and compensation practitioners, such as boards. In many classical principal-agent models (such as Grossman and Hart (1983) and the discrete-time version of Holmstrom and Milgrom (1987)), the optimal contract cannot be solved for in closed form, which poses difficulties for real-life contract design. Moreover, the incentive scheme is contingent upon many specific features of the contracting situation, such as the agent's utility function and noise distribution. This dependence presents further challenges for practitioners, since these parameters likely vary considerably across settings, but many are difficult for the principal to measure and use to guide the contract design. However, observed contracts are typically quite simple and do not depend on the above features.[4] For example, Prendergast's (2002) review of the evidence finds that incentives show little correlation with risk. Our paper offers a simple potential explanation – these details in fact do not matter, and so the contract is robust to such parametric uncertainty.

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 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.[5] Lands-

[4] Murphy's (1999) survey finds that compensation typically comprises cash, bonuses, stock and options. Even mildly complex compensation instruments such as indexed options are rare.

[5] With separable utility, it is straightforward to show that the constraints bind: the principal should offer the least risky contract that induces 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

berger and Meilijson (1994) is an economic theory that uses this notion. We also 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.[6]

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. Holmstrom and Milgrom (1987) show 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; Schaettler and Sung (1993) derive sufficient conditions for the first-order approach to be valid in a larger class of principal-agent problems, of which Holmstrom and Milgrom (1987) is a special case. While these three papers are in continuous time, Hellwig and Schmidt (2002) show that linearity can be achieved in a discrete time multiperiod model, 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. By contrast, 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 Holmstrom and Milgrom 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.

A number of other papers have studied aspects of the principal-agent problem aside from linearity. Matthews (2001) examines the effect of limited liability and monotonicity constraints. DeMarzo and Fishman (2007), He (2008), Sannikov (2008) and Tchistyi (2006) extend the principal-agent model to a dynamic setting. Hall and Murphy (2002), Dittmann and Maug (2007), and Dittmann, Maug and Spalt (2008) analyze its impli-

---

of relative dispersion to prove that constraints bind.

[6]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.

cations for the optimal mix of stock and options. 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 demonstrate that tractability and detail independence obtain in discrete time, where the noise is observed before the agent's action in each period, and in continuous time when these events are simultaneous. 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. Appendix A reviews the mathematical techniques required for the proofs, which are in Appendix B. While the paper considers a unidimensional action and signal, the Online Appendix shows that the contract is robust to multidimensional actions and signals, and contains further extensions and mathematical proofs.

2 The Core Model

Section 2.1 introduces the concept of detail-independence in a multiperiod, discrete time model. In Section 2.2, we show that the optimal contract retains the same functional form in continuous time. Section 2.3 considers a specific application of the model to CEO compensation, where the contract can be implemented using the firm's securities.

2.1 Discrete Time

We consider a T-period model; its key parameters are summarized in Table 1. In each period t, the agent takes an unobservable action a_t. 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.) $\overline{a}$ is the maximum feasible effort level; in Section 3.2 we allow for the maximum productive effort level to be below $\overline{a}$. After the action is taken, a verifiable signal

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

is publicly observed. We assume that noises $\eta_1, ..., \eta_T$ are independent with interval support $\left(\underline{\eta_t}, \overline{\eta_t}\right)$, where the bounds may be infinite, and that $\eta_2, ..., \eta_t$ have log-concave densities.[7] We require no other distributional assumption for η_t; in particular, it need not be Gaussian.

It is unclear from intuition whether it is more realistic to assume that η_t occurs before or after a_t. We thus adopt the assumption that maximizes tractability – that the agent observes η_t before taking a_t in each period. Indeed, we show that this assumption leads to detail-independent contracts. While our assumption is also made in Harris and Raviv (1979), Laffont and Tirole (1986) and Baker (1992), most models feature noise occurring after the action.[8] 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.[9]

In period T, the principal pays the agent cash of c.[10] The agent's utility function is

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

g represents the cost of effort, which is increasing and weakly convex. u is the utility function and v is the felicity 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, $u(x)=x$ denotes additively separable preferences; $v(c)=\ln c$ generates multiplicative preferences, which we will later consider in more detail when applying the model to CEOs. 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. Holmstrom and Milgrom (1987) assume $v(c)=c$ and $u(x)=-e^{-\alpha x}$, i.e. a pecuniary cost of effort and exponential utility. We only assume that the utility

[7] 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.

[8] In a number of corporate finance papers, the state of nature is also revealed before the action is taken. This is because the optimal action depends on the state of nature, and so it is necessary for the state to be realized before the action. (Dewatripont and Tirole (1994) is an example). In models where the optimal action is independent of noise, such as our core case, the common assumption is for noise to follow the action. In Section 3.1 we allow the optimal action to depend on the noise.

[9] The contract is unchanged if noise follows the final action a_T, as long as this noise is realized after the agent receives his contracted payment.

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

function exhibits nonincreasing absolute risk aversion (NIARA), which is sufficient to rule out randomized contracts.

Assumption 1 *(Nonincreasing absolute risk aversion) We assume that* $-u''(x)/u'(x)$ *is nonincreasing in* x*. This is equivalent to* $u'(u^{-1}(x))$ *being weakly convex in* x*.*

Many commonly used utility functions (e.g. constant absolute risk aversion $u(x) = -e^{-\alpha x}$ and constant relative risk aversion $u(x) = x^{1-\gamma}/(1-\gamma)$) exhibit NIARA.

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, although this assumption could be weakened.[11]

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. We will prove that messages play no role and that optimal contracts will be deterministic. 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.
4. The agent sends the principal a message M_1.
5. The agent exerts effort a_1.
6. The signal $r_1 = \eta_1 + a_1$ is publicly observed.
7. Steps (3)-(6) are repeated for $t = 2, ..., T$.
8. The principal pays the agent $\widetilde{c}(r_1, ..., r_T, M_1, ..., M_T)$.

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.[12] 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 $\{\eta_t^*\}_{t=1,..,T}$: by the revelation principle, we can

[11] 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. $\inf_{c,(a_t)} v(c) - \sum_t g(a_t) \leq \sum_t g'(a^*)\left(\underline{\eta_t} + a_t^*\right) + K$.

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

restrict our analysis to mechanisms that induce truth-telling. 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 timet 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 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 message M_t^* given $(\eta_1, ..., \eta_t)$. The program is given below:

Program 1 *Call a_t^* is the target effort level in period t. The principal's problem is to choose a contract $\widetilde{c}(r_1, ..., r_t, M_1, ..., M_T)$, and a recommended $\mathcal{F}_t-$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:

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

The solution to Program 1 is given by Theorem 1 below:

Theorem 1 *(Optimal contract, discrete time). The optimal contract is given by:*

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

where K is a constant that makes the participation constraint bind ($\mathrm{E}\left[u\left(\sum_t g'\left(a_t^*\right) r_t + K - \sum_t g\left(a_t^*\right)\right)\right] = \underline{u}$). The functional form (7) is independent of the utility function u, the reservation utility $\underline{u}$, and the distribution of the noise η. These details 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 optimal contract is given by:

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

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 that conveys the essence of the result for deterministic contracts, using first-order conditions. 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,

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

which integrates to

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

for some constant K.

Proceeding by induction, we now show that, if the result holds for T, it also holds for $T+1$. Let $V\left(r_1, ..., r_{T+1}\right) \equiv v\left(c\left(r_1, ..., r_{T+1}\right)\right)$ 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:

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

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

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

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

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

Defining

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

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

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

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

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

The associated pay is $c = v^{-1}\left(\sum_{t=1}^{T+1} g'\left(a_t^*\right) r_t + K\right)$, as in (7). ■

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 equation (8). Even with time-varying a_t^*, the optimal contract in Theorem 1 can be derived in closed form, which contrasts with the complexity of many classical contracting models.

The timing assumption is key to achieving tractability, and its effect can be seen by examining the heuristic proof. In the final period $T+1$, η_{T+1} is known and so we can remove the expectations operator and in turn the utility function u from equation (6), to obtain (11) and then (12). (These removals would not be possible if η_{T+1} followed a_{T+1}.) 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 separate from 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 functional form of the $t = 1, ..., T$ problems.[13] In short, our timing assumption allows us to reduce the multi-period problem into a succession of one-period problems, each of which can be solved separately.

In addition to deriving the incentive scheme in closed form, Theorem 1 also clarifies the parameters that do and do not matter for optimal contracts. The utility function u, the reservation utility $\underline{u}$, and the distribution of the noise η are all irrelevant. The contract's slope only depends 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. For brevity, we call such a contract "detail-independent." This term aims to highlight that certain details of the contracting situation do not matter; it does not imply that all parameters are irrelevant.

If $v(c) = c$ (i.e. the cost of effort is pecuniary) as assumed by Holmstrom and Milgrom (1987), the contract is linear – regardless of u. This result extends the linear contracts of Holmstrom and Milgrom to settings that do not require exponential utility, Gaussian noise or continuous time. Another common specification is $v(c) = \ln c$, in which case the optimal contract is log-linear. This is considered in more detail in Section 2.3.

We now explain the intuition behind the contract's detail-independence. Inserting the optimal contract (7) into the utility function (2) gives the agent's time t maximization problem as

$$\max_{a_t} \mathrm{E}_t \left[u \left(\begin{array}{c} g'\left(a^*_1\right)\left(a^*_1 + \eta_1\right) - g\left(a^*_1\right) + ... + g'\left(a^*_t\right)\left(a_t + \eta_t\right) - g\left(a_t\right) \\ +... + g'\left(a^*_T\right)\left(a_T + \widetilde{\eta}_T\right) - g\left(a_T\right) + K \end{array} \right) \right]. \qquad (14)$$

The agent's choice of a_t affects his utility only through its effect on the term $\psi_t =$

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

$g'(a_t^*)(a_t + \eta_t) - g'(a_t)$. This term is independent of u, K (and thus $\underline{u}$) and all of the other noise outcomes η_s, $s \neq t$. More explicitly, the agent's first-order condition at time t is

$$\begin{aligned} \mathrm{E}_t\left[u'(\psi_1 + ... + \psi_T + K)\frac{\partial \psi_t}{\partial a_t}\right] &= 0 \\ \implies \mathrm{E}_t\left[u'(\psi_1 + ... + \psi_T + K)\right](g'(a_t^*) - g'(a_t)) &= 0. \end{aligned}$$

The specific functional form of u, K, and the terms $\psi_1, ..., \psi_T$ (and thus η_s, $s \neq t$) only determine the *magnitude* of the increment in utility, $u'(\cdot)$, that results from optimizing ψ_t. However, this magnitude is irrelevant – the only important property is that it is always positive, and this is guaranteed by the monotonicity of u. Hence, regardless of $u'(\cdot)$, the agent will wish to choose the a_t that maximizes ψ_t (i.e. set $a_t = a_t^*$). Simply put, since u is monotonic, it is maximized by maximizing its argument, regardless of its functional form. Moreover, ψ_t enters the argument additively and is independent of K and the other noise realizations. Hence, even though the agent faces residual uncertainty as he does not know $\widetilde{\eta}_{t+1}, ..., \widetilde{\eta}_T$, these noise outcomes do not matter. The irrelevance of unknown noise realizations also means that our contract (7) is incentive compatible even if we reverse our timing assumption and instead allow η_t to follow a_t in each period, as shown in Appendix C. However, we would not be able to prove that contract (7) is optimal (see the above discussion).

Relatedly, the agent's current effort choice is not distorted by past noise realizations. In Mirrlees (1974), the contract involves punishing the agent severely if final output is below a certain threshold: therefore, if he observes that interim output is high (because of high past noise), he will reduce future effort. Holmstrom and Milgrom (1987) assume exponential utility and a pecuniary cost of effort to remove the "wealth effects" caused by past noise realizations. Here, past noise is irrelevant regardless of the parameters.

In sum, regardless of u', $\underline{u}$ or η, the agent's marginal felicity from increasing a_t is $\frac{\partial V}{\partial r_t} = g'(a_t^*)$, and so always equals the marginal cost of effort at the target action level. The final step is to explain how a contract constant in felicity $V(r)$ translates into a contract in terms of cash $c(r)$. If the agent exhibits diminishing marginal felicity (i.e. v is concave), the marginal felicity of a dollar $v'(c)$ falls if past noise outcomes or reservation utility $\underline{u}$ (and thus K) are high. To offset this effect and maintain incentives, the agent must be given a greater number of dollars for exerting effort. Indeed, the v^{-1} transformation means that the contract is convex in cash: high past noise raises

$\frac{\partial c}{\partial r_t}$ to exactly offset the lower $v'(c)$ and maintain the marginal felicity from effort at $v'(c)\frac{\partial c}{\partial r_t} = g'(a_t^*)$.[14]

In Holmstrom and Milgrom (1987), as in Grossman and Hart (1983), effort is modeled as the selection of a probability distribution $p(\cdot)$ 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 – hence noise independence automatically obtains. However, this formulation of effort requires exponential utility to remove "wealth effects" and achieve independence of the reservation wage.[15] By modeling effort as an increment to the signal (equation (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 noise independence through our timing assumption. In sum, the combination of the effort and timing specifications achieves independence of both utility u and noise η.

The Appendix proves 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 (Assumption 1), this cost weakly increases with randomization. Thus, both effects of randomization are undesirable, and deterministic contracts are unambiguously optimal.[16] The Appendix also shows that, even though the agent has information on η_t before choosing a_t, there is no need for him to send a message to the principal regarding η_t. 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.

While the assumption of log-concave densities and independence of η_t's is sufficient to prove Theorem 1, we suspect that the Theorem could be further generalized. Indeed, the heuristic proof above suggests that it will hold even if η_2 is not independent of η_1: it is sufficient that η_2 has interval support, given η_1. Similarly, NIARA may not be

[14] If the cost of effort is pecuniary ($v(c) = c$), $v^{-1}(c) = 1$ and so no transformation is needed. Since both the costs and benefits of effort are in monetary terms, high past noise K diminishes both components of the trade-off equally. Thus, incentives are unchanged even with a linear contract.

[15] Specifically, the utility function is $\sum_j u(v(c_j) - w - c(p, \theta_j))\, p_j$ where w is the reservation wage and the summation is across states θ_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 $-u(-w)\sum_j u(v(c_j) - c(p, \theta_j))\, p_j$. Then the optimal contract under a reservation wage of w is obtained by adding w in all states to the optimal contract under a reservation wage of 0, so the slope is independent of w. This property will not hold with non-exponential utility.

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

necessary to rule out stochastic contracts: even if the marginal cost of effort falls with randomization, this effect may be outweighed by the inefficient risk-sharing. We leave such generalizations for future research.

Remark 1 *(Risk averse principal). The proof of Theorem 1 gives an extension to the case of a risk-averse principal. Suppose that the principal wants to minimize* $\mathrm{E}[f(c)]$*, where* f *is an increasing function, rather than* $\mathrm{E}[c]$*. Then, the above contract is optimal if* $u\left(v\left(f^{-1}(\cdot)\right) - \sum_t g\left(a_t^*\right)\right)$ *is concave. This holds, loosely speaking, the principal is not too risk-averse.*

Remark 2 *(Deterministic contracts, interior actions). The proof of Theorem 1 shows that, if* $a_t^* < \overline{a}\ \forall\ t$ *and the analysis is restricted to deterministic contracts, (7) is the only incentive compatible contract (the value of* K *being the only degree of freedom for the principal). This proof does not require log-concavity of* $\eta_2, ..., \eta_T$ *nor that* u *satisfies NIARA. These assumptions are only required for the full proof (with randomized contracts or the possibility that* $a_t^* = \overline{a}$ *for some* t*), where there exist other incentive compatible contracts, to show that the contract in (7) is least costly.*

2.2 Continuous Time

This section shows that the contract continues to be detail-independent in continuous time, where actions and noise occur simultaneously. Hence, the timing assumption in Section 2.1 leads to consistency between the discrete and continuous time settings. Intuition offers limited guidance as to whether it is more realistic to assume that noise occurs before or after the action in discrete time. 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.

2.2.1 Optimal Contract

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

$$r_t = \int_0^t a_s ds + \eta_t. \tag{15}$$

η_t can be any adapted process, independent of (a_s), for $0 \leq s \leq t$.[17] The agent's utility function is:

$$\mathrm{E}\left[u\left(v(c) - \int_0^T g(a_t)\, dt\right)\right]. \tag{16}$$

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 (16). The optimal contract is detail-independent and of the same form as Theorem 1.

Theorem 2 *(Optimal contract, continuous time). The optimal contract is given by:*

$$c = v^{-1}\left(\int_0^T g'(a_t^*)\, dr_t + K\right), \tag{17}$$

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

In particular, if the target action is time-independent ($a_t^* = a^*\ \forall\ t$), defining the total signal to be $r = \int_0^T r_t dt$, the optimal contract is given by:

$$c = v^{-1}\left(g'(a^*)\, r + K\right). \tag{18}$$

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 Application to CEO Incentives

The optimal incentive scheme in Theorems 1 and 2 can be implemented by any performance-sensitive contract and for any informative signal r. For rank-and-file employees, the stock price is unlikely to be a valuable signal since it will be only weakly affected by their action a. Since CEOs are contracted to maximize shareholder value, and are able to affect it significantly, the firm's equity return is the natural choice of signal r. Therefore, the optimal contract can be implemented using the firm's securities. This section demonstrates this implementation and a number of implications that follow from our

[17] One example is $\eta_t = \int_0^t \sigma_s dZ_s + m_t$, where Z_t is a standard Brownian motion, and σ_t and m_t are deterministic.

tractable contract form. For clarity, we use the discrete time model with $T = 1$ and drop the time subscript.

The baseline firm value is S and the end-of-period stock price P_1 is given by

$$P_1 = Se^a (1 + \varepsilon), \tag{19}$$

where ε is mean-zero noise bounded below by $\underline{\varepsilon} > -1$ with open interval support. Owing to rational expectations, the market anticipates that the CEO will take action $a = a^*$ and so the initial stock price is

$$P_0 = \mathrm{E}[P_1] = Se^{a^*}.$$

Let $R = P_1/P_0$ denote the firm's gross return between periods 0 and 1; $r = \ln R$ is the log return. We thus have

$$r = a + \eta, \tag{20}$$

where noise $\eta = \ln(1 + \varepsilon) - a^*$.

We have previously shown that specifying $v(c) = c$ achieves linear contracts, regardless of $u(\cdot)$. Here, we consider another common assumption of $v(c) = \ln c$, so that the CEO's utility function (2) now becomes, up to a monotonic (logarithmic) transformation:

$$\mathrm{E}\left[U\left(ce^{-g(a)}\right)\right] \geq \underline{U}, \tag{21}$$

where $u(x) \equiv U(e^x)$ and $\underline{U} \equiv \ln \underline{u}$ is the CEO's reservation utility. Utility is now multiplicative in effort and cash 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.[18] Note that we retain the general utility function $U(\cdot)$.

The optimal contract is given by Proposition 1 below:

Proposition 1 *(Optimal CEO contract, one-period model). The optimal contract is given by:*

$$c = kR^{g'(a^*)}, \tag{22}$$

[18] For example, these preferences ensure that the incentive share of total pay is constant across firms, a robust feature of the data. They also give rise an equilibrium pay that, as in Gabaix and Landier (2008), is consistent with the scaling of pay with firm size.

where $R = P_1/P_0$ is the gross firm return and k is a constant that makes the participation constraint bind ($\mathrm{E}\left[U\left(kR^{g'(a^*)}e^{-g(a^*)}\right)\right] = \underline{U}$*).*

Proof. This Proposition is a direct application of Theorem 1 with $v(c) = \ln c$ and $T = 1$. The CEO's utility is

$$U\left(ce^{-g(a)}\right) = u\left(\ln c - g(a)\right),$$

and so the optimal slope is $g'(a^*)$ as before. The optimal contract is $c = v^{-1}\left(g'(a^*)r + K\right) = \exp\left(g'(a^*)\ln R + K\right) = kR^{g'(a^*)}$ where $k = e^K$. ■

As before, the optimal contract (22) is detail-independent, as its shape $R^{g'(a^*)}$ depends only on the unit cost of effort $g'(a^*)$. Noise and the utility function affect only the specific value of k. In Appendix D, we prove that (22) is also optimal in continuous time. Moreover, it has a simple practical implementation if firm returns follow a continuous-time diffusion between period 0 and 1. For simplicity of exposition, we normalize the firm's expected return to zero. At time 0, the CEO is given an "Incentive Account" of value $\mathrm{E}[c]$, of which a fraction $g'(a^*)$ is invested in the firm's stock and the remainder in cash. This portfolio is continuously rebalanced between periods 0 and 1, so that the fraction of wealth invested in the stock remains constant at $g'(a^*)$. The CEO's final wealth therefore becomes (22).

The Incentive Account resembles the restricted stock or long-vesting options commonly awarded in practice. One frequent justification for long-term compensation is to reduce the CEO's incentives to engage in manipulation or myopic behavior. Under this motivation, the optimal vesting point should be time-dependent, i.e. the vesting period should be sufficiently long to allow the effects of short-term actions to be reversed. By contrast, in this model, restricting securities is valuable to maintain effort incentives in a dynamic setting. The optimal vesting point is thus state-dependent: "cashing out" should only be allowed when the fraction of the CEO's wealth in the firm increases sufficiently above $g'(a^*)$, which, in turn, only occurs if the stock price rises. Indeed, Fahlenbrach and Stulz (2008) find that decreases in CEO ownership typically occur after good performance.[19]

The contract can also be implemented by granting the CEO options rather than

[19] This state dependence is similar to the performance vesting that is widespread in the U.K. and is becoming increasingly common in the U.S. It is also featured in the "Bonus Bank" advocated by Stern Stewart, where the amount of the bonus that the executive can withdraw depends on the total bonuses accumulated in the bank.

stock. One option has the same local incentive effect as Δ shares, where Δ is the option's delta. However, if the stock price falls, Δ will decline, reducing the CEO's incentives. This may justify the repricing of out-of-the-money options frequently interpreted as rent extraction (see, e.g., Bebchuk and Fried (2004)).[20] More generally, the model suggests that stock may be preferred to options as an incentive device, since their deltas are constant and less rebalancing is required. Hall and Murphy (2002) and Dittmann and Maug (2007) also advocate stock over options; their arguments focus on the differential risk-incentive trade-off rather than the requirement to maintain a minimum sensitivity.[21]

If the CEO has any initial wealth, $g'(a^*)$ now denotes the fraction of new and existing wealth invested in the firm. Indeed, entrepreneurs and managers of leveraged buyout firms frequently have to invest their existing wealth in their firm. Contract (22) requires the remainder of the CEO's wealth to be invested in cash, but in reality a significant proportion is likely to be in market-sensitive assets such as their own house and non-firm securities. This may justify not indexing the CEO's wealth invested in the firm. One frequent argument against non-indexed stock and options is that they reward the CEO for luck when the market appreciates. However, such non-indexation achieves efficient rebalancing: when the market rises, the CEO's holdings of firm securities appreciate along with his non-firm wealth, so that the fraction of wealth invested in firm securities remains above $g'(a^*)$.

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").

[20] See Saly (1994) and Acharya, John and Sundaram (2000) for other justifications of repricing from an optimal contracting perspective.

[21] Dittmann, Maug and Spalt (2007) show that options can be justified if the CEO is loss-averse and has a low reference wage.

3.1 Contingent Target Actions

Let $A_t(\eta_t)$ denote the "action function", which defines the target action for each noise realization. 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 of the proof that randomized contracts are inferior in Theorem 1, we now restrict the analysis to deterministic contracts. 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 optimal contract is the following. 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 value $\underline{c}$. Otherwise he is paid $C(\widehat{\eta}_1, ..., \widehat{\eta}_T)$, where*

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

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

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, the agent 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 therefore:

$$\eta \in \arg\max_{\widehat{\eta}} V\left(\widehat{\eta}\right) - g\left(A\left(\widehat{\eta}\right) + \widehat{\eta} - \eta\right), \tag{24}$$

with first-order condition

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

Integrating gives the indirect felicity function

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

for constants η_* and k. The associated pay is given by (23). ■

The contract in Theorem 3 is still detail-independent, as its functional form does not depend on $u\left(\cdot\right)$ nor on the distribution of η.[22] However, it is somewhat more complex than the contracts in Section 2, as it involves calculating an integral. In the particular case where $A\left(\eta\right) = a^* \ \forall\ \eta$, Theorem 3 reduces to Theorem 1.

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

$$C\left(\eta_1, ..., \eta_T\right) = v^{-1}\left(k + \sum_{t=1}^{T} g\left(A_t\left(\eta\right)\right) + \int_{\eta_*}^{\eta_t} g'\left(A_t\left(x\right)\right) \frac{R_2\left(A_t\left(x\right), x\right)}{R_1\left(A_t\left(x\right), x\right)} dx\right), \tag{25}$$

where 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 $R\left(a, \eta\right) = R\left(A\left(\widehat{\eta}\right), \widehat{\eta}\right)$. Taking the derivative at $\widehat{\eta} = \eta$ yields $R_1 \partial a / \partial \widehat{\eta} = R_1 A'\left(\eta\right) + R_2$. The agent solves $\max_{\widehat{\eta}} V\left(\widehat{\eta}\right) - g\left(a\left(\widehat{\eta}\right)\right)$, with first-order condition $V'\left(\eta\right) - g'\left(A\left(\eta\right)\right) \partial a / \partial \widehat{\eta} = 0$. Substituting for $\partial a / \partial \widehat{\eta}$ from above and integrating yields equation (25).

[22] Even though (23) features an integral over the support of η, it does not involve the distribution of η.

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

$$S\,\mathrm{E}\left[b\left(\min\left(A\left(\widetilde{\eta}\right),\overline{\overline{a}}\right),\widetilde{\eta}\right)\right]-\mathrm{E}\left[v^{-1}\left(V\left(\widetilde{\eta}\right)\right)\right]. \tag{26}$$

The second term is the expected cost of compensation. It captures both the direct disutility from exerting effort $A(\eta)$, as well as the risk imposed by the incentive contract required to implement $A(\eta)$. The first term captures the benefit 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 $S\,\mathrm{E}\left[b\left(\min\left(A\left(\widetilde{\eta}\right),\overline{\overline{a}}\right),\widetilde{\eta}\right)\right]$ where $b(\cdot)$ is the benefit function of effort and $\overline{\overline{a}}$ is the maximum productive effort level. 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 (as in Section 2.3), $b(a,\eta)=e^{a+\eta}$ and so effort affects firm value multiplicatively. If $\overline{\overline{a}}=\overline{a}$ and $b(a,\eta)=a+\eta$, then the signal r is the principal's objective function (gross of pay).

The $\min\left(A\left(\widetilde{\eta}\right),\overline{\overline{a}}\right)$ 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. 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, described in detail in Appendix E. Clearly, the principal will never wish to implement $a>\overline{\overline{a}}$.

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

Theorem 4 *(Optimality of maximum productive effort). Assume that* $\sup_{\left(\underline{a},\overline{a}\right)}g''$ *and* $\sup_x\overline{F}(x)/f(x)$ *are finite*[23]*, where f is the probability density function of η, and $\overline{F}$ is*

[23] *A sufficient condition for $\sup_x\overline{F}(x)/f(x)$ to be finite 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.*

the complementary cumulative distribution (i.e. $\overline{F}(x) = \Pr(\eta \geq x)$*). Define*

$$\Lambda = \left[\left(1 + \frac{u'(\alpha)}{u'(\beta)} \right) \left(\sup_{(\underline{a}, \overline{\overline{a}})} g'' \right) \left(\sup \frac{\overline{F}}{f} \right) + g'\left(\overline{\overline{a}}\right) \right] \left(v^{-1}\right)' \left(\beta + g\left(\overline{\overline{a}}\right)\right), \qquad (27)$$

where

$$\alpha \equiv u^{-1}(\underline{u}) - \left(\overline{\eta} - \underline{\eta}\right) g'\left(\overline{\overline{a}}\right) \text{ and } \beta \equiv u^{-1}(\underline{u}) + \left(\overline{\eta} - \underline{\eta}\right) g'\left(\overline{\overline{a}}\right). \qquad (28)$$

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*.*

Proof. See Appendix. ■

The intuition is as follows. The benefits of effort are increasing in firm size and also depend on the noise outcome (via the function $b(a, \widetilde{\eta})$). If the firm is sufficiently large ($S > S_*$), the benefits of effort outweigh the costs (summarized by Λ) for all noise outcomes, and so dominate the trade-off. Therefore, maximum productive effort is optimal ($A(\eta) = \overline{\overline{a}}\ \forall\ \eta$.)

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 small noises, $\frac{u'(\alpha)}{u'(\beta)} - 1$ is proportional to the absolute risk aversion of u (when $u \simeq \underline{u}$). 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 ($\inf_{a,\eta} \frac{\partial b}{\partial a}(a, \eta)$).

Considering Theorem 4 from another angle, for every firm size S, there is a range of cost functions g, noise distributions f, and utility/felicity functions u and v (summarized by Λ) for which these details do not matter – maximum effort is optimal for all functions within this range. The larger S is, the wider this range; for very high values of S, $A(\eta) = \overline{\overline{a}}$ is optimal for all plausible cost parameters and so the target effort level is also detail-independent. Then, combined with the results of Section 2, the optimal contract is detail-independent 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.

We conjecture that a "maximum effort principle" holds under more general conditions than the above. For instance, it is likely that it holds 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 $A\left(\eta\right)>\overline{\overline{a}}$ in our setting, but rather 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 Affine 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 $S<S_{*}$, when the cost function is affine – i.e. $g\left(a\right)=\theta a+d$ for $\theta>0$.[24]

Proposition 2 *(Optimal contract with affine cost of effort) Let $g\left(a\right)=\theta a+d$ for $\theta>0$. The optimal contract is given by:*

$$c=v^{-1}\left(\theta r+K\right), \tag{29}$$

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

$$A\left(\eta\right)\in\arg\max_{a\leq\overline{\overline{a}}}Sb\left(a,\eta\right)-v^{-1}\left(\theta\left(a+\eta\right)+d+K\right) \tag{30}$$

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

Proof. From Theorem 3, if the agent announces η, he should receive a felicity of $V\left(\eta\right)=g\left(A\left(\eta\right)\right)+\int_{\eta_{*}}^{\eta}\theta dx+k=\theta\left(A\left(\eta\right)+\eta\right)+K$. Since $r=A\left(\eta\right)+\eta$ on the equilibrium path, a contract $c=v^{-1}\left(\theta r+K\right)$ will implement $A\left(\eta\right)$. To find the optimal action, the principal's problem is:

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

[24]Note that the linearity of $g(a)$ is still compatible with $u\left(v\left(c\right)-g\left(a\right)\right)$ being strictly concave in (c,a).

which is solved by pointwise maximization, as in (30). ■

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). 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 $\max_a v\left(c\left(r\right)\right) - g\left(a\right)$, i.e. $\max_a \theta\left(\eta + a\right) + K - \left(\theta a + b\right)$, 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, $A\left(\eta\right)$.

4 Conclusion

This paper has identified and analyzed a class of situations in which the optimal contract is both tractable and detail-independent. The contract can be solved in closed form, and its functional form is independent of a number of parameters of the agency problem, such as the utility function, reservation utility, and the distribution of noise. In particular, 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 independence 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. Moreover, if the firm is sufficiently large, the target effort level is itself detail-independent: 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 are second-order in a large firm, so maximum effort is efficient.

The model extends the tractable contracts of Holmstrom and Milgrom (1987) to settings that do not require exponential utility nor continuous time. Moreover, it may explain why real-life incentive schemes are typically simple, even though utility functions and noise distributions vary considerably across settings and are difficult to observe – simply put, these details do not matter.

Our paper suggests several avenues for future research. The Holmstrom and Milgrom (1987) framework has proven valuable in many areas of applied contract theory owing to its tractability. Our tractable contracts may similarly be used in any context that embeds a contracting situation. While we considered the specific application of executive compensation, other possibilities include bank regulation, team production, or insurance. 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 addition, while our model has relaxed a number of assumptions required for tractability, our setup continues to require a number of restrictions. These are mostly technical rather than economic. For example, we have assumed a continuum of actions rather than a discrete set; our multiperiod model requires independent noises with log-concave density functions; and our extension to noise-dependent target actions assumes an open action space and a maximal 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
C	Consumption (wage) provided by contract
$\mathcal{C}$	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, $V \equiv v(C)$

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. In addition, if $\mathrm{E}[X] = \mathrm{E}[Y]$, relative dispersion implies that X second-order stochastically dominates Y (see Lemma 2 below). Hence, it is a stronger concept than second-order stochastic dominance.

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:*

$$\mathrm{E}[f(X)] \geq \mathrm{E}[f(Y)] \Rightarrow \mathrm{E}[h(f(X))] \geq \mathrm{E}[h(f(Y))]$$

$$\mathrm{E}[f(X)] \leq \mathrm{E}[f(Y)] \Rightarrow \mathrm{E}[\phi(f(X))] \leq \mathrm{E}[\phi(f(Y))].$$

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.

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

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

$\widehat{u}$ has 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 start with 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:*

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

We take limits "from below," because we will often apply the concept of the subderivative at the maximum feasible effort level $\overline{a}$. This notion will prove useful since $f'_-(x)$ is well-defined for all functions f (with perhaps infinite values). 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. Its proof is 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 $a < b$ 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.[25]

The following Lemma is the chain rule for subderivatives and is proved in the Online Appendix.

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. For instance, the following Lemma is proved in the Online Appendix.

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. In particular, $\mathrm{E}[f(\widetilde{\eta})]$ denotes the expectation over all realizations of $\widetilde{\eta}$. $\mathrm{E}\left[\widetilde{f}(\widetilde{\eta})\right]$ 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. The

[25] 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 $f(-1) > f(1)$.

second part proves the theorem considering only deterministic contracts and assuming that the target effort levels are in the interior of the action space $\mathcal{A}$. This case requires weaker assumptions (see Remark 2), and is easy to generalize. 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, and so the third part can be skipped at a first reading.

1). Redundancy of Messages

Let $\mathbf{r}$ denote the vector $(r_1, ..., r_T)$ and define $\boldsymbol{\eta}$ and $\mathbf{a}$ analogously. Define $\mathbf{g}(\mathbf{a}) = g(a_1) + ... + g(a_T)$. Under the revelation principle, we can restrict the analysis to mechanisms that induce the agent to truthfully report the noise $\boldsymbol{\eta}$. 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 $\mathbf{r}$. The incentive compatibility (IC) constraint is that the agent exerts effort $\mathbf{a}$ and reports $\widehat{\boldsymbol{\eta}} = \boldsymbol{\eta}$:

$$\forall \eta, \forall \widehat{\eta}, \forall \mathbf{a},\ \mathrm{E}\left[u\left(\widetilde{V}_M(\boldsymbol{\eta}+\mathbf{a}, \widehat{\boldsymbol{\eta}}) - \mathbf{g}(\mathbf{a})\right)\right] \leq \mathrm{E}\left[u\left(\widetilde{V}_M(\boldsymbol{\eta}+\mathbf{a}^*, \boldsymbol{\eta}) - \mathbf{g}(\mathbf{a}^*)\right)\right]. \quad (31)$$

The principal's problem is to minimize expected pay $\mathrm{E}\left[v^{-1}\left(\widetilde{V}_M(\widetilde{\boldsymbol{\eta}}+\mathbf{a}^*, \widetilde{\boldsymbol{\eta}})\right)\right]$, subject to the IC constraint (31), and the agent's individual rationality (IR) constraint

$$\mathrm{E}\left[u\left(\widetilde{V}_M(\widetilde{\boldsymbol{\eta}}+\mathbf{a}^*, \widetilde{\boldsymbol{\eta}}) - \mathbf{g}(\mathbf{a}^*)\right)\right] \geq \underline{u}. \quad (32)$$

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

$$\widetilde{V}(\mathbf{r}) = \widetilde{V}_M(\mathbf{r}, \mathbf{r} - \mathbf{a}^*). \quad (33)$$

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

$$\forall \eta, \forall a,\ \mathrm{E}\left[u\left(\widetilde{V}(\mathbf{r}) - g(\mathbf{a})\right)\right] \leq \mathrm{E}\left[u\left(\widetilde{V}(\mathbf{r}^*) - g(\mathbf{a}^*)\right)\right], \quad (34)$$

$$\mathrm{E}\left[u\left(\widetilde{V}(\mathbf{r}^*) - g(\mathbf{a}^*)\right)\right] \geq \underline{u}. \quad (35)$$

If the agent reports $\widehat{\boldsymbol{\eta}} \neq \boldsymbol{\eta}$, the principal will expect to observe signal $\widehat{\boldsymbol{\eta}}+\mathbf{a}^*$. He must therefore take action $\mathbf{a}$ such that $\widehat{\boldsymbol{\eta}}+\mathbf{a}^*=\boldsymbol{\eta}+\mathbf{a}$. Substituting $\widehat{\boldsymbol{\eta}} = \boldsymbol{\eta}+\mathbf{a}-\mathbf{a}^*$ into (31) and (32) indeed yields (34) and (35) 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 definition (33). Hence, the new contract $\widetilde{V}(\mathbf{r})$ induces incentive compatibility and participation at the same cost as the initial contract $\widetilde{V}_M(\mathbf{r},\boldsymbol{\eta})$ with messages, and so messages are not useful. The intuition is that $\mathbf{a}^*$ is always exerted, so the principal can already infer $\boldsymbol{\eta}$ from the signal $\mathbf{r}$ without requiring messages.

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

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:

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

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

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

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

$$g(a^*) - g(a^* + r' - r) \leq V(r) - V(r') \leq g(a^* + r - r') - g(a^*). \tag{36}$$

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

$$\frac{g(a^*) - g(a^* + r' - r)}{r - r'} \leq \frac{V(r) - V(r')}{r - r'} \leq \frac{g(a^* + r - r') - g(a^*)}{r - r'}. \tag{37}$$

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 (37) converge to $g'(a^*)$. Therefore, the left derivative $V'_{left}(r)$ exists, and equals $g'(a^*)$. Second, consider $r < r'$. Dividing (36) 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,

$$V'(r) = g'(a^*). \tag{38}$$

Since r has interval support[26], we can integrate to obtain:

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

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 $g'(a_t^*)$. Consider the set of (potentially randomized) contracts $\widetilde{V}(r_1, ..., r_T)$ such that (i) $\mathrm{E}\left[u\left(\widetilde{V}(\widetilde{r}_1, ..., \widetilde{r}_T)\right)\right] \geq \underline{u}$; (ii) $\forall\ t = 1...T$,*

$$\frac{d}{d\varepsilon_-}\mathrm{E}\left[u\left(\widetilde{V}(\widetilde{r}_1, ..., \widetilde{r}_t+\varepsilon, ..., \widetilde{r}_T)\right) \mid \widetilde{r}_1, .., \widetilde{r}_t\right]_{|\varepsilon=0} \geq g'(a_t^*)\,\mathrm{E}\left[u'\left(\widetilde{V}(\widetilde{r}_1, ..., \widetilde{r}_t, ..., \widetilde{r}_T)\right) \mid \widetilde{r}_1, .., \widetilde{r}_t\right] \tag{39}$$

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

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

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

We first consider the case of deterministic contracts, and then show that randomized contracts lead to a higher cost. We use the notation $\mathrm{E}_t[\cdot] = \mathrm{E}[\cdot \mid \widetilde{r}_1, ..., \widetilde{r}_t]$ 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 $a_t^* = \overline{a}$.

[26] 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.

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

The contract is here assumed to be deterministic. Equation (39) reads: $\frac{d}{d\varepsilon}_{-} u\left(V\left(r+\varepsilon\right)\right) \geq g'\left(a_1^*\right) u\left(V\left(r\right)\right)$. By Lemma 5, applied to $h = u^{-1}$, this yields:

$$V'_{-}\left(r\right) \geq g'\left(a^*\right). \tag{40}$$

It is intuitive that (40) should bind, as this minimizes the variability in the agent's pay and thus constitutes efficient risk-sharing. We now prove mathematically that this is indeed the case; to simplify exposition, we normalize $g\left(a^*\right) = 0$ without loss of generality.[27] If constraint (40) 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 (40) is weaklier costlier.

By assumption (iii) in Statement Σ_1, V is nondecreasing. We can therefore can apply Lemma 4 to equation (40), 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:

$$\mathrm{E}\left[u\left(V\left(\widetilde{r}\right)\right)\right] \geq \underline{u} = \mathrm{E}\left[u\left(V^0\left(\widetilde{r}\right)\right)\right]. \tag{41}$$

Applying Lemma 2 to the convex function $\phi \circ u^{-1}$ and inequality (41), we have:

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

i.e. $\mathrm{E}\left[\phi\left(V\left(\widetilde{r}\right)\right)\right] \geq \mathrm{E}\left[\phi\left(V^0\left(\widetilde{r}\right)\right)\right]$. 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 Remark 1, the extension to a risk-averse principal. Suppose that the principal wants to minimize $\mathrm{E}\left[f\left(c\right)\right]$, where f is an increasing and concave function, rather than $\mathrm{E}\left[c\right]$. Then, the above contract is optimal if $f \circ v^{-1} \circ u^{-1}$ is convex, i.e. $u \circ v \circ f^{-1}$ is concave. This requires f to be "not too concave," i.e. the agent to be not too risk-averse.

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

[27] 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 $u\left(x - g\left(a\right)\right) = u^{new}\left(x - g^{new}\left(a\right)\right)$.

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:

$$\widehat{u}(x) = \mathrm{E}\left[u\left(x + g'\left(a^*_{T+1}\right)\widetilde{r}_{T+1}\right)\right]. \tag{42}$$

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 $k\left(\widetilde{r}_1, ..., \widetilde{r}_T\right)$ as the solution to equation (43) below:

$$\widehat{u}\left(k\left(\widetilde{r}_1, ..., \widetilde{r}_T\right)\right) = \mathrm{E}_T\left[u\left(V\left(\widetilde{r}_1, ..., \widetilde{r}_{T+1}\right)\right)\right]. \tag{43}$$

k thus 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 (43) over $\widetilde{r}_1, ..., \widetilde{r}_T$ yields:

$$\mathrm{E}\left[\widehat{u}\left(k\left(\widetilde{r}_1, ..., \widetilde{r}_T\right)\right)\right] = \mathrm{E}\left[u\left(V\left(\widetilde{r}_1, ..., \widetilde{r}_{T+1}\right)\right)\right] \geq \underline{u}, \tag{44}$$

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

Since equation (39) 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:

$$\frac{dV}{dr_{T+1}}_-\left(r_1, ..., r_{T+1}\right) \geq g'\left(a^*_{T+1}\right). \tag{45}$$

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 $k(\widetilde{r}_1,...,\widetilde{r}_T)+g'\left(a^*_{T+1}\right)\widetilde{r}_{T+1}$.

Using (42), we can rewrite equation (43) as

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

Since u exhibits NIARA (see Assumption 1), $u'\circ u^{-1}$ is a convex function. We can thus apply Lemma 2 to yield:

$$\begin{aligned}\mathrm{E}_T\left[u'\circ u^{-1}\circ u\left(V(\widetilde{r}_1,...,\widetilde{r}_{T+1})\right)\right]&\geq \mathrm{E}_T\left[u'\circ u^{-1}\circ u\left(k(\widetilde{r}_1,...,\widetilde{r}_T)+g'\left(a^*_{T+1}\right)\widetilde{r}_{T+1}\right)\right], \text{ i.e.}\\ \mathrm{E}_T\left[u'\left(V(\widetilde{r}_1,...,\widetilde{r}_{T+1})\right)\right]&\geq \mathrm{E}_T\left[\widehat{u}'\left(k(\widetilde{r}_1,...,\widetilde{r}_T)\right)\right].\end{aligned} \tag{46}$$

Applying definition (43) to the left-hand side of Condition (ii) for $T+1$ yields, with $t=1...T$,

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

Taking expectations of equation (46) at time t and substituting into the right-hand side of the above equation yields:

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

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

$$\widehat{\phi}(x)\equiv \mathrm{E}\left[\phi\left(x+g'(a_{T+1})\,\widetilde{r}_{T+1}\right)\right]. \tag{47}$$

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:

$$\mathcal{C}_k=\mathrm{E}\left[\widehat{\phi}\left(k(\widetilde{r}_1,...,\widetilde{r}_T)\right)\right]\geq \mathcal{C}_{V^0}=\mathrm{E}\left[\phi\left(\sum_{t=1}^{T+1}g'(a^*_t)\,\widetilde{r}_t+K\right)\right]. \tag{48}$$

Using equation (47) yields:

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

Finally, we compare the cost of contract $k\left(r_1,...,r_T\right) + g'\left(a_{T+1}\right)\widetilde{r}_{T+1}$ to the cost of the original contract $V\left(r_1,...,r_{T+1}\right)$. Since equation (43) 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 (48). 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 $\widetilde{V}\left(r_1,...,r_T\right)$ and define the "certainty equivalent" contract $\overline{V}$ by:

$$u\left(\overline{V}\left(r_1,...,r_T\right)\right) \equiv \mathrm{E}_T\left[u\left(\widetilde{V}\left(r_1,...,r_T\right)\right)\right]. \tag{49}$$

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 (32). Hence, Condition (i) holds. Also, it is clear 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 (49) and the function $u' \circ u^{-1}$ (which is convex by Assumption 1) yields: $u'\left(\overline{V}\left(r_1,...,r_T\right)\right) \leq \mathrm{E}_T\left[u'\left(\widetilde{V}\left(r_1,...,r_T\right)\right)\right]$. We apply this to $r_t = \widetilde{r}_t$ for $t = 1...T$ and taking expectations to obtain

$$\mathrm{E}_t\left[u'\left(\widetilde{V}\left(\widetilde{r}_1,...,\widetilde{r}_T\right)\right)\right] \geq \mathrm{E}_t\left[u'\left(\overline{V}\left(\widetilde{r}_1,...,\widetilde{r}_T\right)\right)\right]. \tag{50}$$

Applying definition (49) to the left-hand side of (39) 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 (50) 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 (49) 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

$$\overline{u}\left(x\right) \equiv u\left(x - \sum_{s=1}^{T} g\left(a_s^*\right)\right) \tag{51a}$$

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.

$$0 \in \arg\max_{\varepsilon} \mathrm{E}_t\, u\left(\widetilde{V}\left(\widetilde{r}_1, ..., \widetilde{r}_t + \varepsilon, ..., \widetilde{r}_T\right) - g\left(a_t^* + \varepsilon\right) - \sum_{s=1...T, s\neq t} g\left(a_s^*\right)\right). \tag{52}$$

We note that, for a function $f\left(\varepsilon\right)$, $0 \in \arg\max_{\varepsilon} f\left(\varepsilon\right)$ implies that for all $\varepsilon < 0$, $\left(f\left(0\right) - f\left(\varepsilon\right)\right)/\left(-\varepsilon\right) \geq 0$, hence, taking the $\liminf_{y\uparrow 0}$, $\frac{d}{d\varepsilon_-} f'_-\left(\varepsilon\right)_{|\varepsilon=0} \geq 0$. Call $X\left(\varepsilon\right)$ the argument of u in equation (52). Applying this to (52), we find: $\frac{d}{d\varepsilon_-} \mathrm{E}_t\, u\left(X\left(\varepsilon\right)\right)_{|\varepsilon=0} \geq 0$.

Using Lemma 5, we find $\mathrm{E}_t\left[u'\left(X\left(0\right)\right)\left(\frac{d}{d\varepsilon_-} X\left(\varepsilon\right)_{|\varepsilon=0}\right)\right] \geq 0$. Using Lemma 6,

$\frac{d}{d\varepsilon_-}X(\varepsilon)_{|\varepsilon=0} = \frac{d}{d\varepsilon_-}\widetilde{V}(\widetilde{r}_1, ..., \widetilde{r}_t + \varepsilon, ..., \widetilde{r}_T) - g'(a_t^*)$, hence we get:

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

Using again Lemma 5, this can be rewritten:

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

i.e., using the notation (51a),

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

Therefore, Condition (ii) of Statement Σ_T holds.

Finally, we verify Condition (iii). Apply (52) to signal r_t and deviation $\varepsilon < 0$. We obtain:

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

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 $V^0 = \sum_{t=1}^T g'(a_t^*)\,r_t + K$. Moreover, it is clear that V^0 satisfies the global IC conditions in equation (52). Thus, V^0 is the cheapest contract that satisfies the global IC constraint.

Proof of Theorem 2

We shall use the following purely mathematical Lemma, proven (using Malliavin calculus) 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, $T \geq 0$, $X = \int_0^T \alpha_t dZ_t$, and $Y =$*

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

Lemma 7 is intuitive: since $\beta_t \geq \alpha_t \geq 0$, it makes sense that Y is more volatile than X.

We define $v_T = v(c)$. To derive the IC condition, we use the methodology introduced by Sannikov (2008). By the martingale representation theorem, we can write: $v_T = \int_0^T \theta_t dr_t + v_0$ for some constant v_0 and an adapted process θ_t. The IC constraint is

$$a_t^* \in \arg\max \theta_t a_t dt - g(a_t)\, dt,$$

i.e. $\theta_t = g'(a_t^*)$ 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: since $\theta_t = g'(a_t^*)$, we have

$$v_T = \int_0^T g'(a_t^*)\, dr_t + v_0 = g'(a_t^*)\, r_T + v_0,$$

where v_0 is a constant that satisfies $\mathrm{E}[u(g'(a_t^*)\, r_T + v_0)] = \underline{u}$.

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

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

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.

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: $\forall \eta, \forall \widehat{\eta}$,

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

Defining

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

the truth-telling constraint (53) can be rewritten,

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

Rewriting this inequality interchanging η and $\widehat{\eta}$ and combining with the original inequality (54) yields:

$$\forall \eta, \forall \widehat{\eta} : g(A(\widehat{\eta})) - g(A(\widehat{\eta}) + \widehat{\eta} - \eta) \leq \psi(\eta) - \psi(\widehat{\eta}) \leq g(A(\eta) + \eta - \widehat{\eta}) - g(A(\eta)). \tag{55}$$

Consider a point η where A is continuous and take $\widehat{\eta} < \eta$. Dividing (55) 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 (55) by $\eta - \widehat{\eta} < 0$, and taking the limit $\widehat{\eta} \downarrow \eta$ yields $\psi'_{right}(\eta) = g'(A(\eta))$. Hence,

$$\psi'(\eta) = g'(A(\eta)), \tag{56}$$

at all points η where A is continuous.

Equation (56) holds only almost everywhere, since we have only assumed that A is almost everywhere continuous. To complete the proof, we need 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 (55) implies:

$$\begin{aligned}
|\psi(\eta) - \psi(\widehat{\eta})| &\leq \max\left\{|g(A(\widehat{\eta})) - g(A(\widehat{\eta}) + \widehat{\eta} - \eta)|, g(A(\eta) + \eta - \widehat{\eta}) - g(A(\eta))\right\} \\
&\leq |\eta - \widehat{\eta}| (\sup g')_I .
\end{aligned}$$

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

$$V(\eta) = g(A(\eta)) + \int_{\eta_*}^{\eta} g'(A(x))\, dx + k \tag{57}$$

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:

$$\eta_{T+1} \in \arg\max_{\widehat{\eta}_{T+1}} V(\eta_1, ..., \eta_T, \widehat{\eta}_{T+1}) - g(A(\widehat{\eta}_{T+1}) + \widehat{\eta}_{T+1} - \eta_{T+1}) - \sum_{t=1}^{T} g(a_t^*). \tag{58}$$

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

$$V(\eta_1, ..., \eta_T, \widehat{\eta}_{T+1}) = W_{T+1}(\widehat{\eta}_{T+1}) + k(\eta_1, ..., \eta_T), \tag{59}$$

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 $\widehat{\eta} = \eta_t\ \forall t = 1...T$, 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

$$\widehat{u}(x) = \mathrm{E}\left[u\left(x + W_{T+1}(\widetilde{\eta}_{T+1})\right)\right], \tag{60}$$

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 (57), 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 $\mathrm{E}\left[v^{-1}(V(\widetilde{\eta}))\right]$ will weakly decrease.

Let $\mathcal{C}(A)$ denote the expected cost of implementing $A(\eta)$, i.e. $\mathcal{C}(A) = \mathrm{E}\left[v^{-1}(C(\eta))\right]$ 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 (27) 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 (26) 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:

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

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

C 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 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 $T = 1$ and give a heuristic proof that assumes validity of the first-order approach; the rigorous proof is similar to Appendix B.

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. The first-order condition is

$$\mathrm{E}\left[u'\left(v\left(c\left(a+\eta\right)\right) - g\left(a\right)\right)\left(v'\left(c\left(a+\eta\right)\right)c'\left(a^*+\eta\right) - g'\left(a\right)\right)\right] = 0. \tag{61}$$

Under the contract in Theorem 1, $v\left(c\left(r\right)\right) = g'\left(a^*\right)r$ which yields:

$$\mathrm{E}\left[u'\left(v\left(c\left(a+\eta\right)\right) - g\left(a\right)\right)\right]\left(g'\left(a^*\right) - g'\left(a\right)\right) = 0$$

Since $u'\left(\cdot\right) > 0$, we must have $g'\left(a^*\right) - g'\left(a\right) = 0$, i.e. $a = a^*$ as required. However, we can no longer prove from (61) that the contract in Theorem 1 is optimal. Since η is unknown, $u'\left(v\left(c\left(a+\eta\right)\right) - g\left(a\right)\right)$ is not a constant and thus cannot be taken outside the expectation term.

D Application to CEO Incentives, Continuous Time

In the core model, we derived the optimal CEO contract in discrete time (Proposition 1). Here, we show that the contract is also optimal in discrete time.

The baseline firm value is S and the end-of-period stock price P_1 is given by

$$P_1 = S \exp\left(\int_0^1 a_s ds - \frac{\sigma_s^2}{2} ds + \sigma_s dZ_s\right),$$

where Z_t is a standard Brownian motion and σ_t is a deterministic (possibly non-constant) volatility process. The principal wishes to implement action a^* at each instant. By rational expectations, the initial stock price is $P_0 = \mathrm{E}[P_1] = Se^{a^*}$ and the log return up to time t is

$$r_t = \ln P_t / P_0 = \int_0^t \left[(a_s - a^*)\, ds + \sigma_s dZ_s - \frac{\sigma_s^2 ds}{2}\right].$$

By rational expectations, a_t increases the drift of r_t by $a_t - a^*$, which on the equilibrium path will be zero (up to the Jensen's inequality term $\int_0^t \sigma_s^2 ds/2$).

As in Section 2.3, we use $v(c) = \ln c$ so that the CEO has multiplicative preferences. His utility function becomes:

$$\mathrm{E}\left[U\left(c \exp\left(-\int_0^T g(a_t)\, dt\right)\right)\right]. \tag{62}$$

The optimal contract is given below.

Proposition 3 *(Optimal CEO contract, continuous time). The optimal contract is given by:*

$$c = kR^{g'(a^*)}. \tag{63}$$

where $R = P_1/P_0$ is the gross firm return and k is a constant that makes the participation constraint bind ($\mathrm{E}\left[U\left(kR^{g'(a^*)}e^{-g(a^*)T}\right)\right] = \underline{U}$).

Proof. This Proposition is a direct application of Theorem 2 with $u(x) \equiv U(e^x)$ and $v(c) = \ln c$. The CEO's utility is:

$$U\left(c \exp\left(-\int_0^T g(a_t)\, dt\right)\right) = u\left(\ln c - \int_0^T g(a_t)\, dt\right).$$

In addition,

$$r_1 = \ln R = \int_0^1 (a_s - a^*)\, ds + \sigma_s dZ_s - \frac{\sigma_s^2 ds}{2}$$

The optimal contract is thus

$$c = v^{-1}\left(g'\left(a^*\right) r + K\right) = \exp\left(g'(a^*) \ln R + K\right) = e^K R^{g'(a^*)}.$$

■

E A Microfoundation for the Principal's Objective

We offer a microfoundation for the principal's objective function (26). Suppose that the agent can take two actions, a "fundamental" action $a^F \in (\underline{a}, \overline{\overline{a}}]$ 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 $v\left(c\right) - \left[g^F\left(a\right) + G\left(m\right)\right]$, where g, G are increasing and convex, $G\left(0\right) = 0$, and $G'\left(0\right) \geq g'\left(\overline{\overline{a}}\right)$. The final assumption means that manipulation is costlier than fundamental effort.

We define $a = a^F + m$ and the cost function $g\left(a\right) = \min_{a^f, M}\left\{g^F\left(a\right) + G\left(m\right) \mid a^F + m = a\right\}$, so that $g\left(a\right) = g^F\left(a\right)$ for $a \in (\underline{a}, \overline{\overline{a}}]$ and $g\left(a\right) = g^F\left(a\right) + g\left(m - a\right)$ 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 (26).

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 $A\left(\eta\right)$ and for the market to set price $P_1 = e^r$.

References

[1] Acharya, Viral, Kose John and Rangarajan Sundaram (2000): "On the Optimality of Resetting Executive Stock Options." *Journal of Financial Economics* 57, 65-101

[2] Arnott, Richard and Joseph Stiglitz (1988) "Randomization with Asymmetric Information." *RAND Journal of Economics* 19, 344-362

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

[4] Bebchuk, Lucian and Jesse Fried (2004): "Pay Without Performance: The Unfulfilled Promise of Executive Compensation." Cambridge, MA: Harvard University Press

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

[6] DeMarzo, Peter and Michael Fishman (2007): "Agency and Optimal Investment Dynamics." *Review of Financial Studies* 20, 151-188.

[7] Dewatripont, Mathias and Jean Tirole (1994): "A Theory of Debt and Equity: Diversity of Securities and Manager-Shareholder Congruence." *Quarterly Journal of Economics* 109, 1027-1054

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

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

[10] Edmans, Alex, Xavier Gabaix and Augustin Landier (2008): "A Multiplicative Model of CEO Incentives in Market Equilibrium." Working Paper, University of Pennsylvania

[11] Fahlenbrach, Rüdiger and René Stulz (2008): "Managerial Ownership Dynamics and Firm Value." Working Paper, Ohio State University

[12] Gabaix, Xavier and Augustin Landier (2008): "Why Has CEO Pay Increased So Much?" *Quarterly Journal of Economics*, 123, 2008, 49-100

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

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

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

[16] He, Zhiguo (2008): "Optimal Executive Compensation when Firm Size Follows Geometric Brownian Motion." *Review of Financial Studies*, forthcoming

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

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

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

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

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

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

[23] Matthews, Steven A. (2001): "Renegotiation of Moral Hazard Contracts under Limited Liability and Monotonicity ," *Journal of Economic Theory*, 97, 1-29

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

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

[26] 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

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

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

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

[30] Rudin, Walter (1987): *Real and Complex Analysis*, Third Edition, McGraw-Hill

[31] Saly, P. Jane (1994): "Repricing Executive Stock Options in a Down Market." *Journal of Accounting and Economics* 18, 325-356

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

[33] 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

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

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

[36] Tchistyi, Alexei (2006): "Security Design with Correlated Hidden Cash Flows: The Optimality of Performance Pricing." Working Paper, New York University

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