DISCUSSION PAPER SERIES

No. 7578

TRACTABILITY IN INCENTIVE CONTRACTING

Alex Edmans and Xavier Gabaix

FINANCIAL ECONOMICS

Centre for Economic Policy Research

www.cepr.org

Available online at: **www.cepr.org/pubs/dps/DP7578.asp**

ISSN 0265-8003

TRACTABILITY IN INCENTIVE CONTRACTING

Alex Edmans, Wharton School of Management
Xavier Gabaix, NYU Stern, NBER and CEPR

Discussion Paper No. 7578
November 2009

Centre for Economic Policy Research
53–56 Gt Sutton St, London EC1V 0DG, UK
Tel: (44 20) 7183 8801, Fax: (44 20) 7183 8820
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 7578

November 2009

ABSTRACT

Tractability in Incentive Contracting

This paper identifies a class of multiperiod agency problems in which the optimal contract is tractable (attainable in closed form). By modeling the noise before the action in each period, we force the contract to provide sufficient incentives state-by-state, rather than merely on average. This tightly constrains the set of admissible contracts and allows for a simple solution to the contracting problem. Our results continue to hold in continuous time, where noise and actions are simultaneous. We thus extend the tractable contracts of Holmstrom and Milgrom (1987) to settings that do not require exponential utility, a pecuniary cost of effort, Gaussian noise or continuous time. The contract's functional form is independent of the noise distribution. Moreover, if the cost of effort is pecuniary (multiplicative), the contract is linear (log-linear) in output and its slope is independent of the noise distribution, utility function and reservation utility. In a two-stage contracting game, the optimal target action depends on the costs and benefits of the environment, but is independent of the noise realization.

JEL Classification: D2, D3, G34 and J3
Keywords: closed forms, contract theory, dispersive order, executive compensation, incentives, principal-agent problem and subderivative

Alex Edmans
The Wharton School
University of Pennsylvania
2428 Steinberg Hall - Dietrich Hall
3620 Locust Walk, Philadelphia
PA 19104-6367
USA

Email: aedmans@wharton.upenn.edu

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=164005

Xavier Gabaix
Finance Department
Stern School of Business
New York University
44 West 4th Street, 9-190
New York, NY 10012,
USA

Email: xgabaix@stern.nyu.edu

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=139904

Submitted 19 November 2009

For helpful comments, we thank three anonymous referees, the Editor (Stephen Morris), Andy Abel, Frankin Allen, Heski Bar-Isaac, Patrick Cheridito, Peter DeMarzo, Ingolf Dittmann, Florian Ederer, Chris Evans, Itay Goldstein, Gary Gorton, Narayana Kocherlakota, Ernst Maug, Holger Mueller, Christine Parlour, David Pearce, Canice Prendergast, Michael Roberts, Tomasz Sadzik, Yuliy Sannikov, Nick Souleles, Rob Stambaugh, Luke Taylor, Rob Tumarkin, Bilge Yilmaz, and seminar participants at AEA, Five Star, Chicago, Columbia, Harvard-MIT Organizational Economics, Minneapolis Fed, Northwestern, NYU, Princeton, Richmond Fed, Stanford, Toulouse, WFA, Wharton and Wisconsin. We thank Andrei Savotchkine and 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. This paper was formerly circulated under the title “Tractability and Detail-Neutrality in Incentive Contracting.”

1 Introduction

The principal-agent problem is central to many economic settings, such as employment contracts, insurance, taxation and regulation. A vast literature analyzing this problem has found that it is typically difficult to solve: even in simple settings, the optimal contract can be highly complex (see, e.g., Grossman and Hart (1983)). The first-order approach is often invalid, requiring the use of more intricate techniques. Even if an optimal contract can be derived, it is often not attainable in closed form, which reduces tractability – a particularly important feature in applied theory models.

Against this backdrop, Holmstrom and Milgrom (1987, "HM") made a major breakthrough by showing that the optimal contract is linear in profits under certain conditions. Their result has since been widely used by applied theorists to justify assuming a linear contract, which leads to substantial tractability. However, HM emphasized that their result only holds under exponential utility, a pecuniary cost of effort, Gaussian noise, and continuous time. These assumptions may not hold in a number of situations – for example, there is ample evidence of decreasing absolute risk aversion, and many effort decisions do not involve a monetary expenditure (e.g. exerting effort rather than shirking, or forgoing private benefits). In addition, in certain settings, the modeler may wish to use discrete time or binary noise for simplicity.

Can tractable contracts be achieved in broader settings? When allowing for alternative utility functions or noise distributions, do these details affect the form of the optimal contract? What factors do and do not matter for the incentive scheme? These questions are the focus of our paper. We consider a discrete-time, multiperiod model where the agent consumes only in the final period. We first solve for the cheapest contract that implements a given, but possibly time-varying, path of target effort levels. The optimal incentive scheme is tractable, i.e. attainable in closed form. The key source of tractability is our timing assumption that, in each period, the agent first observes noise and then exerts effort, before observing the noise in the next period. This is similar to theories in which the agent observes total cash flow before deciding how much to divert (e.g. Lacker and Weinberg (1989), DeMarzo and Sannikov (2006), DeMarzo and Fishman (2007) and Biais et al. (2007)). Since the agent knows the noise realization when taking his action, incentive compatibility requires the agent's marginal incentives to be sufficient state-by-state (i.e. for every possible noise outcome), which tightly constrains the set of admissible contracts. By contrast, if the action were taken before the noise, incentive compatibility would only pin down marginal incentives on average. There are many possible contracts that induce incentive compatibility on average, and the problem is complex as the principal must solve for the cheapest contract out of this continuum. Note that the timing assumption does not change the fact that the agent faces uncertainty when deciding his effort level since each action, except the final one, continues to be followed by noise. Even in a one-period model, the agent faces risk after signing the contract.

The analysis demonstrates what features of the environment do and do not matter for the optimal implementation contract. The contract's functional form is independent of the agent's noise distribution and reservation utility, i.e. it can be written without references to these parameters. The functional form depends only on how the agent trades off the benefits of cash against the cost of providing effort. Moreover, the contract's slope, as well as its functional form, is independent of the agent's utility function, reservation utility and noise distribution in two cases. First, if the cost of effort is pecuniary as in HM (i.e. can be expressed as a subtraction to cash pay), the incentive scheme is linear in output regardless of these parameters, even if the cost of effort is itself non-linear. Second, if the agent's preferences are multiplicative in cash and effort, the contract is independent of utility and log-linear, i.e. the percentage change in pay is linear in output. This robustness contrasts with many classical principal-agent models (e.g. Grossman and Hart (1983)), where even the implementation contract is contingent upon many specific features of the contracting situation. This poses practical difficulties, as some of the important determinants are difficult for the principal to observe and thus use to guide the contract, such as the noise distribution and agent's utility function. Our results suggest that, under some specifications, the implementation contract is robust to such parametric uncertainty.

Closed-form solutions allow the economic implications of a contract to be transparent. We consider a application to CEO incentives to demonstrate the implications that can flow from a tractable contract structure. For CEOs, the appropriate output measure is the percentage stock return, and multiplicative preferences are theoretically motivated by Edmans, Gabaix and Landier (2009). The percentage change in pay is thus linear in the percentage change in firm value, i.e. the relevant measure of incentives is the elasticity of pay with respect to firm value. This analysis provides a theoretical justification for using elasticities to measure incentives, a metric previously advocated by Murphy (1999) on empirical grounds.

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. Using recent advances in continuous-time contracting (Sannikov (2008)), we then show that the contract retains the same form in a continuous-time model where noise and effort occur simultaneously. This consistency suggests that, if underlying reality is continuous time, it is best approximated in discrete time by modeling noise before effort in each period.

We next allow the target effort path to depend on the noise realizations. The optimal contract now depends on messages sent by the agent regarding the noise. However, it remains tractable, for a given "action function" that links the observed noise to the principal's recommended effort level. We then solve for the optimal action function chosen by the principal. In classical agency models, the chosen action is the result of a trade-off between the benefits of effort (which are increasing in firm size) and its costs (direct disutility plus the risk imposed by incentives, which are of similar order of magnitude to the agent's wage). We show that, if the output under the agent's control is sufficiently large compared to his salary (e.g. the agent is a CEO who affects total firm value), these trade-off considerations disappear: the benefits of effort swamp the costs. Thus, maximum effort is optimal, regardless of the noise outcome.

The "maximum effort principle"[1], when applicable, significantly increases tractability, since it removes the need to solve the trade-off required to derive the optimal effort level when it is interior. Indeed, jointly deriving the optimal effort level and the efficient contract that implements it can be highly complex. Thus, many contracting papers focus exclusively (e.g. Dittmann and Maug (2007) and Dittmann, Maug and Spalt (2009)) or predominantly (e.g. Grossman and Hart (1983), Lacker and Weinberg (1989), Biais et al. (2009), He (2009a, 2009b)) on implementing a fixed target effort level; see also the overview of the literature in Chapters 4 and 8 in Laffont and Martimort (2002). Our result rationalizes this approach: if maximum effort is always efficient, the problem of deriving optimal effort has a simple solution – there is no trade-off to be simultaneously tackled and the analysis can focus on the cheapest contract to implement this effort level.

Finally, we allow the principal to choose the maximum productive effort level depending on the costs and benefits of the environment. We extend the model to a two-stage game. In the first stage, the principal chooses the maximum productive effort level, e.g. by selecting the size of the plant. In the second stage, the contract is played out as before – the principal wishes the agent to run the plant (whatever its size) with maximum efficiency. As in standard models, the effort level set in the first stage is typically decreasing in the agent's risk aversion, cost of effort and noise dispersion. Thus, our setup allows for contracts that are simple (since the maximum effort principle applies in the second stage and so solving for a trade-off is not required) yet still respond to the costs and benefits of the environment and thus generate comparative static predictions.

In sum, our analysis generates a set of sufficient conditions to obtain tractable contracts. For the implementation contract to be tractable, modeling the action after the noise is sufficient; for the full contract that also solves for the optimal effort level, ex-post actions plus a high benefit of effort are sufficient – in turn, large firm size is sufficient (although not necessary) for the latter. These sufficient conditions are quite different from the HM assumptions of exponential utility, a pecuniary cost of effort, Gaussian noise, and continuous time, and so may be satisfied in many settings in which the HM assumptions do not hold and tractability was previously believed to be unattainable.

We achieve simple contracts in other settings than HM due to a different modeling setup. In a dynamic setting, high prior period outcomes increase the agent's wealth and distort the current period decision through two "wealth effects." First, higher wealth affects the agent's current risk aversion and thus effort choice. HM assume exponential utility to remove this effect. Second, higher wealth reduces the agent's marginal utility of money; if the marginal cost of effort is unchanged, the agent has fewer incentives to exert effort. This problem occurs with any risk-averse utility function, including exponential utility. HM assume that the cost of effort is pecuniary, so that it also declines when wealth increases. HM require these two assumptions

[1] 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.

to remove the intertemporal link between periods and allow the multiperiod problem to collapse into a succession of identical static problems. Even the single-period problem remains potentially complex, since many contracts satisfy the incentive compatibility condition on average. HM address this by giving the agent substantial freedom – rather than simply selecting the mean return of the firm, he has control over the probabilities of N different states of nature.[2] This freedom simplifies the contracting problem by reducing the set of allowable contracts. However, this formulation is more cumbersome since effort is the choice of a probability vector, and is thus relatively seldom used in applied theory models.

We model effort as a scalar that affects the firm's mean return, because this formulation is most commonly used in theoretical applications owing to its simplicity. We instead give the agent freedom by specifying the noise before the action – a choice that is not possible when effort involves the selection of probabilities, since noise unavoidably follows the action. In addition to achieving tractability by forcing the contract to hold state-by-state, the timing assumption also removes the need for exponential utility by allowing the multiperiod model to be solved by backward induction, so that it becomes a succession of single-period problems. In the single-period problem, the noise is observed before the action – thus, the agent's risk aversion is unimportant and exponential utility is not required. A potential intertemporal link remains since high past outcomes, or high current noise, mean that the agent already expects high consumption and thus has a lower incentive to exert effort, if he exhibits diminishing marginal utility. This issue is present in the Mirrlees (1974) contract if the agent can observe past outcomes. Put differently, in the single-period problem, the agent does not face *risk* (as the noise is known) but faces *distortion* (as the noise affects his effort incentives). The optimal contract must address these issues: if the utility function is concave, the contract is convex so that, at high levels of consumption, the agent is awarded a greater number of dollars for exerting effort, to offset the lower marginal utility of each additional dollar. Allowing for convex contracts also allows us to drop the second critical assumption of a pecuniary cost of effort. Even if high wealth reduces the marginal utility of cash but not the marginal cost of effort, incentives are preserved because the contract is steeper at high wealth levels.

In addition to its results, the paper's proofs import and extend some mathematical techniques that are relatively rare in economics and may be of use in future models. We 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. It allows us to avoid the first-order approach, and so may be useful for models where sufficient conditions for the first-order approach cannot be verified.[3] We also use the notion of "relative

[2] This specification refers to the discrete-time version of the HM model, as this is most comparable to our setting. In that version, the contract is linear in accounts, although not linear in profits.

[3] See Rogerson (1985) for sufficient conditions for the first-order approach to be valid under a single signal, and Jewitt (1988) for situations in which the principal can observe multiple signals. Schaettler and Sung (1993)

dispersion" to prove that the incentive compatibility constraints bind, i.e. the principal imposes the minimum slope that induces effort. We show that the binding contract is less dispersed than alternative solutions, constituting efficient risk sharing. A similar argument rules out stochastic contracts, where the payout is a random function of output.[4] We extend a result from Landsberger and Meilijson (1994), who use relative dispersion in another economic setting.

This paper builds on a rich literature on tractable multiperiod agency problems. HM show the optimal contract is linear in profits under exponential utility and a pecuniary cost of effort, if the agent controls only the drift of the process and time is continuous; 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 and a pecuniary cost of effort. In Sung (1995) and Ou-Yang (2003), the agent also controls the diffusion of the process in continuous time. Hellwig and Schmidt (2002) achieve linearity in discrete time, under the additional assumptions that the agent can destroy profits before reporting them to the principal, and that the principal can only observe output in the final period. Our setting allows the principal to observe signals in each period. Mueller (2000) shows that linear contracts are not optimal in HM if the agent can only change the drift at discrete points, even if these points are numerous and so the model closely approximates continuous time.

Our modeling of noise before the action is most similar to models in which the agent can observe total cash flow before deciding how much to divert. Lacker and Weinberg (1989) show that the optimal contract to deter all diversion (the analog of maximum effort) is piecewise linear, regardless of the noise distribution and utility function. Their core result is similar to a specific case of our Theorem 1, restricted to a pecuniary cost of effort and a single period. In DeMarzo and Sannikov (2006), DeMarzo and Fishman (2007) and Biais et al. (2007), the optimal contract is linear because the agent is risk-neutral – therefore, there is no issue with wealth affecting risk aversion (which is always zero) nor the marginal benefit of diversion (which is constant for each dollar diverted). The risk-neutral version of Garrett and Pavan (2009) also predicts linear contracts. Our setting considers risk aversion, where high past output reduces the marginal benefit of effort, thus requiring a convex contract to preserve incentives.

This paper proceeds as follows. In Section 2 we derive tractable contracts in both discrete and continuous time, given a target path of effort levels. Section 3 allows the effort level to depend on the noise realization, derives conditions under which maximum productive effort is optimal for all noise outcomes, and allows the principal to determine this maximum according to the environment. Section 4 concludes. The Appendix contains proofs and other additional materials; further peripheral material is in the Online Appendix.

---

derive sufficient conditions for the first-order approach to be valid in a large class of principal-agent problems, of which HM is a special case.

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

2 The Core Model

2.1 Discrete Time

We consider a T-period model; its key parameters are summarized in Table 1. In each period t, the agent observes noise η_t, takes an unobservable action a_t, and then observes the noise in period $t+1$. The action a_t is broadly defined to encompass any decision that benefits output but is personally costly to the principal. The main interpretation is effort, but it can also refer to rent extraction: low a_t reflects cash flow diversion or the pursuit of private benefits. We assume that noises $\eta_1, ..., \eta_T$ are independent with interval support with interior $\left(\underline{\eta_t}, \overline{\eta_t}\right)$, where the bounds may be infinite, and that $\eta_2, ..., \eta_t$ have log-concave densities.[5] We require no other distributional assumption for η_t; in particular, it need not be Gaussian. The action space $\mathcal{A}$ has interval support, bounded below and above by $\underline{a}$ and $\overline{a}$. We allow for both open and closed action sets and for the bounds to be infinite. After the action is taken, a verifiable signal

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

is publicly observed at the end of each period t.

Insert Table 1 about here

Our assumption that η_t precedes a_t is featured in models in which the agent sees total output before deciding how much to divert (e.g. Lacker and Weinberg (1989), DeMarzo and Fishman (2007), Biais et al. (2007)), or observes the "state of nature" before choosing effort (e.g. Harris and Raviv (1979), Sappington (1983), Baker (1992), and Prendergast (2002)[6]). Note that this timing assumption does not make the agent immune to risk – in every period, except the final one, his action is followed by noise. Even in a one-period model, the agent bears risk as the noise is unknown when he signs the contract. In Section 2.2 we show that the contract has the same functional form in continuous time, where η and a are simultaneous. While the timing assumption extends the model's applicability to a cash flow diversion setting (an application that is not possible if noise follows the action), a limitation is that η cannot be interpreted as measurement error.

[5] 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.

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

In period T, the principal pays the agent cash of c.[7] 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[8] function which denotes the agent's utility from cash; both are increasing and weakly concave. g, u and v are all twice continuously differentiable. We specify functions for both utility and felicity to maximize the generality of the setup. For example, the utility function $\left(ce^{-g(a)}\right)^{1-\gamma}/\left(1-\gamma\right)$ is commonly used in macroeconomics (see e.g. Cooley and Prescott (1995)), which entails $u\left(x\right)=e^{(1-\gamma)x}/\left(1-\gamma\right)$ (with $\gamma>1$ so that u is concave; when $\gamma=1$, the limit is understood as $u\left(x\right)=x$) and $v\left(x\right)=\ln x$. The case $u(x)=x$ denotes additively separable preferences; $v(c)=\ln c$ generates multiplicative preferences. If $v(c)=c$, the cost of effort is expressed as a subtraction to cash pay. This is appropriate if effort represents an opportunity cost of foregoing an alternative income-generating activity (e.g. outside consulting), or involves a financial expenditure. HM assume $u(x)=-e^{-\gamma x}$ and $v(c)=c$.

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

The agent's reservation utility is given by $\underline{u}\in\operatorname{Im}u$, where $\operatorname{Im}u$ is the image of u, i.e. the range of values taken by u. We assume that $\operatorname{Im}v=\mathbb{R}$ so that we can apply the v^{-1} function to any real number.[9] We take an optimal contracting approach that imposes no restrictions on the contracting space available to the principal, so the contract $\widetilde{c}(\cdot)$ can be stochastic, nonlinear in the signals r_t, and depend on messages M_t sent by the agent. By the revelation principle, we can assume that the the space of messages M_t is $\mathbb{R}$ and that the principal wishes to induce truth-telling by the agent. The full timing is as follows:

1. The principal proposes a (possibly stochastic) contract $\widetilde{c}\left(r_1,...,r_T,M_1,...,M_T\right)$.
2. The agent agrees to the contract or receives his reservation utility $\underline{u}$.
3. The agent observes noise η_1, sends the principal a message M_1, then exerts effort a_1.
4. The signal $r_1=\eta_1+a_1$ is publicly observed.

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

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

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

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

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

Throughout most of the paper, we abstract from imperfect commitment problems and focus on a single source of market imperfection: moral hazard. This assumption is common in the dynamic moral hazard literature: see, e.g., Rogerson (1985), HM, Spear and Srivastava (1987), Phelan and Townsend (1991), Biais et al. (2007, 2009). The Online Appendix extends the model to accommodate quits and firings.

As in Grossman and Hart (1983), in this section we fix the path of effort levels that the principal wants to implement at $(a_t^*)_{t=1,..,T}$, where $a_t^* > \underline{a}$ and a_t^* may be time-varying.[10] 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 $(\eta_t)_{t=1,..,T}$. The principal is risk-neutral, and so the optimal contract is the admissible contract with the lowest expected cost $\mathrm{E}\,[\widetilde{c}]$. Section 3 studies the optimal effort level.

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

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

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

subject to the following constraints:

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

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

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

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

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

Theorem 1 below describes our solution to Program 1.[11]

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

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

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

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

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

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

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

$$v'\left(c\left(a_1^*+\eta_1\right)\right) c'\left(a_1^*+\eta_1\right) - g'\left(a_1^*\right) = 0. \tag{10}$$

[11] Theorem 1 characterizes a contract that is optimal, i.e. solves Program 1. Strictly speaking, there exist other optimal contracts which pay the same as (7) on the equilibrium path, but take different values for returns that are not observed on the equilibrium path. Note that the contract in Theorem 1 allows c to be negative. Limited liability could be incorporated, at the cost of additional notational complexity, by imposing a lower bound on $\underline{\eta}$ or adding a fixed constant to the signal.

Therefore, for all r_1,

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

which integrates over η_1 to

$$v\left(c\left(r_1\right)\right) = g'\left(a_1^*\right)r_1 + K \tag{11}$$

for some constant K. Contract (11) must hold for all r_1 that occurs with non-zero probability, i.e. for $r_1 \in \left(a_1^* + \underline{\eta}_1, a_1^* + \overline{\eta}_1\right)$.

We will proceed now by induction on the total number of periods T: 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 IC condition is:

$$a_{T+1}^* \in \arg\max_{a_{T+1}} V\left(r_1, ..., r_T, \eta_{T+1} + a_{T+1}\right) - g\left(a_{T+1}\right) - \sum_{t=1}^{T} g\left(a_t^*\right). \tag{12}$$

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

$$V\left(r_1, ..., r_T, r_{T+1}\right) = g'\left(a_{T+1}^*\right)r_{T+1} + k\left(r_1, ..., r_T\right), \tag{13}$$

where the integration "constant" now depends on the past signals, i.e. $k\left(r_1, ..., r_T\right)$. In turn, $k\left(r_1, ..., r_T\right)$ is 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

$$\widehat{u}\left(x\right) = \mathrm{E}\left[u\left(x + g'\left(a_{T+1}^*\right)r_{T+1} - g\left(a_{T+1}^*\right)\right)\right], \tag{14}$$

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, 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 (11), the contract must satisfy:

$$V\left(r_1, ..., r_T, r_{T+1}\right) = \sum_{t=1}^{T+1} g'\left(a_t^*\right)r_t + K. \tag{15}$$

for (r_t) that occurs with non-zero probability (i.e. $(r_1, ..., r_T) \in \prod_{t=1}^{T} \left(a_t^* + \underline{\eta}_t, a_t^* + \overline{\eta}_t\right)$. The associated pay is $c = v^{-1}\left(\sum_{t=1}^{T+1} g'(a_t^*)\, r_t + K\right)$, as in (7). Conversely, any contract that satisfies (15) is incentive compatible. ■

Theorem 1 yields a closed-form contract for any T and (a_t^*). The Theorem also clarifies the parameters that do and do not matter for the contract's functional form. It depends only on the felicity function v and the cost of effort g, i.e. how the agent trades off the benefits of cash against the costs of providing effort, and is independent of the utility function u, the reservation utility $\underline{u}$, and the distribution of the noise η. Even though these parameters do not affect the contract's functional form, in general they will affect its slope via their impact on the scalar K. However, if $v(c) = c$ (the cost of effort is pecuniary) as assumed by HM, the contract's slope is also independent of u, $\underline{u}$ and η: it is linear, regardless of these parameters. The linear contracts of HM can thus be achieved in settings that do not require exponential utility, Gaussian noise or continuous time. Note that, even if the cost of effort is pecuniary, it remains a general, possibly non-linear function $g(a_t)$.

The origins of the contract's tractability can be seen in the heuristic proof. We first consider $T = 1$. Since η_1 is known, the expectations operator can be removed from (6). u then drops out to yield (9). The specific form of u is irrelevant – all that matters is that it is monotonic, and so it is maximized by maximizing its argument. In particular, exponential utility is not required – the agent's attitude to risk does not matter as η_1 is known. In turn, (9) yields the first-order condition (10), which must hold for every possible realization of η_1, i.e. *state-by-state*. This pins down the slope of the contract: for all η_1, the agent must receive a marginal felicity of $g'(a_1^*)$ for a one unit increment to the signal r_1. The principal's only degree of freedom is the constant K, which is itself pinned down by the participation constraint.

By contrast, if η_1 followed the action, and assuming linear u for simplicity, (10) would be

$$E\left[v'(c(r_1))\, c'(r_1)\right] = g'(a_1^*). \tag{16}$$

This first-order condition only determines the agent's marginal incentives *on average*, rather than state-by-state. There are multiple contracts that will satisfy (10) and implement a_1^*, and the problem is significantly more complex as the principal must solve for the cheapest contract out of this continuum. By giving the agent greater flexibility in the action space (by allowing him to respond to η_1), our timing assumption simplifies the contracting problem by tightly constraining the set of incentive compatible contracts. This is similar to the intuition behind the linear contracts of HM, who give the agent flexibility by granting him control over not just the mean signal, but the probability of each realization. Equation (8) shows that, if the target action (and thus marginal cost of effort) is constant, incentives must be constant time-by-time as well as state-by-state, and so only aggregate performance ($r = \sum_{t=1}^{T} r_t$) matters.

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

The idea of subjecting the agent to a constant incentive pressure is also similar to HM. However, in HM, the constant incentive pressure involves giving the agent a constant increase in cash for an increase in the signal. Here, the agent is given a constant increase in felicity, $v'(c(r_1))\,c'(r_1)$. This generalization allows us to drop the assumption of a pecuniary cost of effort, in which case the contract is non-linear. In the cash flow diversion models of DeMarzo and Sannikov (2006), DeMarzo and Fishman (2007) and Biais et al. (2007), the optimal contract is linear because the agent is risk-neutral. His utility rises by a *constant* amount θ for each dollar diverted, and so the optimal contract must give him a constant share θ of output. Lacker and Weinberg (1989) achieve a (piecewise) linear contract with general utility functions and noise distributions, under a pecuniary cost of effort and for $T = 1$. We extend their result to general T and a non-pecuniary cost of effort.

We now move to $T > 1$. In all periods $t < T$, the agent is now exposed to risk, since he does not know future noise realizations when he chooses a_t. Much like the effect of a high current noise realization, if the agent expects future noise to be high, his incentives to exert effort will be reduced. This would typically require the agent to integrate over future noise realizations when choosing a_t, leading to high complexity. Here the unknown future noise outcomes do not matter, as can be seen in the heuristic proof. Before $T+1$, η_{T+1} is unknown. However, (13) shows that the unknown η_{T+1} enters additively and does not affect the incentive constraints of the $t = 1, ..., T$ problems – regardless of what η_{T+1} turns out to be, the contract must give the agent a marginal felicity of $g'(a_t^*)$ for exerting effort at t.[12] Our timing assumption thus allows us to solve the multiperiod problem via backward induction, reducing it to a succession of one-period problems, each of which can be solved tractably.

[12] This can be most clearly seen in the definition of the new utility function (14), which "absorbs" the $T+1$ period problem.

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

The Appendix proves that, even though the agent privately observes η_t, there is no need for him to communicate it to the principal. Since a_t^* is implemented for all η_t, there is a one-to-one correspondence between r_t and η_t on the equilibrium path. The principal can thus infer η_t from r_t, rendering messages redundant. The Appendix also rules out randomized contracts. There are two effects of randomization. First, it leads to inefficient risk-sharing, for any concave u. Second, changing the reward for effort from a certain payment to a lottery may increase or decrease his effort incentives.[13] We show that with NIARA utility, this second effect is negative. Thus, both effects of randomization are undesirable, and deterministic contracts are unambiguously optimal. The proof makes use of the independence of noises and the log-concavity of $\eta_2, ..., \eta_T$. While these assumptions, combined with NIARA utility, are sufficient to rule out randomized contracts, they may not be necessary. In future research, it would be interesting to explore whether randomized contracts can be ruled out in broader settings.[14]

In addition to allowing for stochastic contracts, the above analysis also allows for $a_t^* = \overline{a}$, under which the IC constraint is an inequality. Therefore, the contract in (7) only provides a lower bound on the contract slope. A sharper-than-necessary contract has a similar effect to a stochastic contract, since it subjects the agent to additional risk. Again, the combination of NIARA and independent and log-concave noises is sufficient rule out such contracts.

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

Proposition 1 *(Optimal deterministic contract, $a_t^* < \overline{a} \ \forall \ t$). Consider only deterministic contracts and $a_t^* < \overline{a} \ \forall \ t$. Relax the assumptions of NIARA utility, independent noises, and*

[13] See Arnott and Stiglitz (1988) for detail on how randomization can sometimes be desirable – if low effort leads to a random payoff, this may induce the agent to induce effort. They derive sufficient conditions under which randomization is suboptimal. Our conditions to guarantee the suboptimality of random contracts generalize their results to broader agency problems (their setting focuses on insurance).

[14] For instance, consider $T = 2$. We only require that $\widehat{u}(x)$ as defined in (43) exhibits NIARA. The concavity of η_2 is sufficient, but unnecessary for this. Separately, if NIARA is violated, the marginal cost of effort falls with randomization. However, this effect may be outweighed by the inefficient risk-sharing, so randomized contracts may still be dominated.

log-concave noises for $\eta_2, ..., \eta_T$. Any incentive-compatible contract takes the form

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

where K is a constant. The optimal deterministic contract features a K that makes the agent's participation constraint bind.

Proof. See Appendix. ■

The following Remark states that the contract's incentive compatibility is robust to the timing assumption. In particular, if noise follows the action in each period, the contract in Theorem 1 continues to implement the target actions – since it provides sufficient incentives state-by-state, it automatically does so on average. However, we can no longer show that it is optimal, since there are many other contracts that provide sufficient incentives on average.

Remark 1 *(Robustness of the contract's incentive compatibility to timing). For any timing of the noise $(\eta_t)_{t=1...T}$ (i.e. regardless of whether it follows or precedes a_t in each period), the contract in Theorem 1 is incentive compatible and implements $(a_t^*)_{t=1,..,T}$. Indeed, given the contract, the agent's utility is:*

$$u\left(\sum_{t=1}^{T} g'\left(a_t^*\right)\left(a_t + \eta_t\right) + K - \sum_{t=1}^{T} g\left(a_t\right)\right),$$

so that, regardless of the timing of $(\eta_t)_{t=1...T}$, the agent maximizes his utility by taking action $a_t = a_t^*$, as it solves $\max_{a_t} g'\left(a_t^*\right) a_t - g\left(a_t\right)$.

Closed-form solutions allow the economic implications of a contract to be transparent. We close this section by considering two specific applications of Theorem 1 to executive compensation, to highlight the implications that can be gleaned from a tractable contract structure. While contract (7) can be implemented for any informative signal r, the firm's log equity return is the natural choice of r for CEOs, since they have a fiduciary duty to maximize shareholders value. When the cost of effort is pecuniary ($v\left(c\right) = c$), Theorem 1 implies that the CEO's dollar pay c is linear in the firm's return r. Hence, the relevant incentives measure is the dollar change in CEO pay for a given percentage change in firm value (i.e. "dollar-percent" incentives), as advocated by Hall and Liebman (1998).[15]

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

$$\mathrm{E}\left[U\left(ce^{-g(a)}\right)\right] \geq \underline{U}, \tag{18}$$

[15] This incentive measure refers to "ex ante" incentives, i.e. how much the CEO's pay will change over the next year if the stock return over the next year increases by one percentage point.

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; Edmans, Gabaix and Landier (2009) show that multiplicative preferences are necessary to generate empirically consistent predictions for the scaling of various measures of CEO incentives with firm size. Thus, the ability to drop the HM assumption of $v(c) = c$ becomes valuable. Applying Theorem 1 with $T = 1$ for simplicity, the optimal contract becomes

$$\ln c = g'(a^*)r + K. \tag{19}$$

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

2.2 Continuous Time

This section shows that the contract has the same tractable form in continuous time, where actions and noise are simultaneous. This consistency suggests that, if reality is continuous time, it is best approximated in discrete time by modeling noise before effort in each period.

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

$$r_t = \int_0^t a_s ds + \eta_t, \tag{20}$$

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

$$\mathrm{E}\left[u\left(v(c) - \int_0^T g(a_t)\,dt\right)\right]. \tag{21}$$

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 (21). The optimal contract is of the same tractable form as Theorem 1.

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

$$c = v^{-1}\left(\int_0^T g'(a_t^*)\,dr_t + K\right), \tag{22}$$

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

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

$$c = v^{-1}\left(g'(a^*)\, r_T + K\right) \tag{23}$$

is optimal.

Proof. See Appendix. ■

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

2.3 Discussion: What is Necessary for Tractable Contracts?

The framework considered thus far shows that tractable implementation contracts can be achieved without requiring exponential utility, a pecuniary cost of effort, continuous time or Gaussian noise. However, it has still imposed a number of restrictions. We now discuss the features that are essential for our contract structure, inessential features that we have already relaxed in extensions, and additional assumptions which may be relaxable in future research.

1. *Timing of noise.* This assumption is central to the intuition of attaining simple contracts as it restricts the principal's flexibility. Remark 1 states that, if a_t precedes η_t, contract (7) still implements $(a_t^*)_{t=1,..,T}$. However, we can no longer show that it is optimal.

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

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

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

5. *Linear signal, $r_t = a_t + \eta_t$.* Remark 2 in Section 3.1 later shows that with general signals $r_t = R(a_t, \eta_t)$, the optimal contract remains tractable and its functional form remains independent of u, $\underline{u}$ and the distribution of η.

6. *Timing of consumption.* The current setup assumes that the agent only consumes at the end of period T. In Edmans, Gabaix, Sadzik and Sannikov (2009), we develop the analog of Theorem 1 where the agent consumes in each period, for the case of $v(c) = \ln c$ and a CRRA utility function. The contract remains tractable.

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

3 The Optimal Effort Level

The analysis has thus far focused on the optimal implementation of a given path of effort levels (a_t^*). In Section 3.1 we allow the target effort level to depend on the current period noise. Section 3.2 derives conditions under which the principal wishes to implement the maximum productive effort level for all noise realizations (the "maximum effort principle"). Section 3.3 allows the principal to choose the maximum productive effort level according to the environment.

3.1 Contingent Target Actions

Let $A_t(\eta_t)$ denote the "action function", which defines the target action for each noise realization. (Thus far, we have assumed $A_t(\eta_t) = a_t^*$.) Since different noises η_t may lead to the same observed signal $r_t = A_t(\eta_t) + \eta_t$, the analysis must consider revelation mechanisms. 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 also assume that the signal $A_t(\eta_t) + \eta_t$ is nondecreasing in η_t: otherwise, as the proof of Proposition 2 shows, the action function cannot be implemented – if a higher noise corresponds to a significantly lower action, the agent would over-report the noise and exert less effort. We make three additional technical assumptions: the action space $\mathcal{A}$ is open, $A_t(\eta_t)$ is bounded within any compact subinterval of η, and $A_t(\eta_t)$ is almost everywhere continuous. The final assumption still allows for a countable number of jumps in $A_t(\eta_t)$. Given the complexity and length of the proof that randomized contracts are inferior in Theorem 1, we now restrict the analysis to deterministic contracts and assume $A_t(\eta_t) < \overline{a}$. We conjecture that the same arguments in that proof continue to apply with a noise-dependent target action.

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

Proposition 2 *(Optimal contract, noise-dependent action). A series of contingent action $(A_t(\eta_t))_{t=1...T}$ can be implemented if and only if for all t, $A_t(\eta_t) + \eta_t$ is nondecreasing in η_t. If* *that condition is verified, the following contract is optimal. For each* t*, after noise* η_t *is realized, the agent communicates a value* $\widehat{\eta}_t$ *to the principal. If the subsequent signal is not* $A_t(\widehat{\eta}_t) + \widehat{\eta}_t$ *in each period, he is paid a very low amount* $\underline{c}$*. Otherwise he is paid* $C(\widehat{\eta}_1, ..., \widehat{\eta}_T)$*, where*

$$C(\eta_1, ..., \eta_T) = v^{-1}\left(\sum_{t=1}^{T} g(A_t(\eta_t)) + \sum_{t=1}^{T}\int_{\eta_*}^{\eta_t} g'(A_t(x))\,dx + K\right), \quad (24)$$

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

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

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

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

The first-order condition is

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

Integrating over η gives the indirect felicity function

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

for constants η_* and K. The associated pay is given by (24). ■

The contract in Proposition 2 remains in closed form and its functional form does not depend on u, $\underline{u}$ nor the distribution of η.[16] However, it is somewhat more complex than the contracts in Section 2, as it involves calculating an integral. In the particular case where $A(\eta) = a^* \ \forall\ \eta$, Proposition 2 reduces to Theorem 1.

Remark 2 *(Extension of Proposition 2 to general signals). Suppose the signal is a general function* $r_t = R(a_t, \eta_t)$*, where* R *is differentiable and has positive derivatives in both arguments,* $R_1(a, \eta)/R_2(a, \eta)$ *is nondecreasing in* a*, and* $R(A_t(\eta_t), \eta_t)$ *is nondecreasing in* η_t*. The same*

[16] Even though (24) features an integral over the support of η, it does not involve the distribution of η.

analysis as in Proposition 2 derives the following contract as optimal:

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

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

The heuristic proof is as follows (setting $T = 1$ and dropping the time subscript). If η is observed and the agent reports $\widehat{\eta} \neq \eta$, he has to take action a such that $R(a, \eta) = R(A(\widehat{\eta}), \widehat{\eta})$. Taking the derivative at $\widehat{\eta} = \eta$ yields $R_1 \partial a / \partial \widehat{\eta} = R_1 A'(\eta) + R_2$. The agent solves $\max_{\widehat{\eta}} V(\widehat{\eta}) - g(a(\widehat{\eta}))$, with first-order condition $V'(\eta) - g'(A(\eta)) \partial a / \partial \widehat{\eta} = 0$. Substituting for $\partial a / \partial \widehat{\eta}$ from above and integrating over η yields (25).

3.2 Maximum Effort Principle

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

$$\max_{\{a(\eta)\}} \int b(a(\eta), \eta) f(\eta) d\eta - C[A]. \tag{26}$$

The first term represents the productivity of effort, where $a(\eta) = \min\left(A(\eta), \overline{\overline{a}}\right)$ and $\overline{\overline{a}} < \overline{a}$ is the maximum productive effort level. The $\min\left(A(\eta), \overline{\overline{a}}\right)$ function conveys the fact that, while the action space may be unbounded ($\overline{a}$ may be infinite), there is a limit to the number of productive activities the agent can undertake to benefit the principal. For example, in a cash flow diversion model, $\overline{\overline{a}}$ reflects zero stealing; in an effort model, there is a limit to the number of hours a day the agent can work while remaining productive. In a project selection model, there is a limit to the number of positive-NPV projects available; $\overline{\overline{a}}$ reflects taking all of these projects while rejecting negative-NPV projects. In addition to being economically realistic, this assumption is useful technically as it prevents the optimal action from being infinite. Actions $a > \overline{\overline{a}}$ do not benefit the principal, but improve the signal: one interpretation is manipulation (see Appendix C for further details). Clearly, the principal will never wish to implement $a > \overline{\overline{a}}$. For brevity, we use "maximum effort" to refer to maximum *productive* effort $\overline{\overline{a}}$. $b(\cdot)$ is the productivity function of effort which is differentiable with respect to $a(\eta)$. $f(\eta)$ is the density of η, assumed to be finite. The second term, $C[A]$, is the expected cost of the contract required to implement $A(\eta)$ (we suppress the dependence on η for brevity).

We assume that g is strictly convex, and that $g \circ (g')^{-1}$ and g' are convex; this assumption is satisfied for many standard cost functions, e.g. $g(a) = Ga^2$ and $g(a) = e^{Ga}$ for $G > 0$. The following Proposition bounds the difference in the costs of the contract implementing maximum effort, and an arbitrary contract:[17]

Proposition 3 *(Bound on difference in costs.) There exists a function* $\lambda\left(\overline{\overline{a}},\eta\right)$ *such that, for all plans* $\{a(\eta)\}$ *where* $\forall\eta, a(\eta)\leq\overline{\overline{a}}$,

$$C\left[\overline{\overline{A}}\right]-C[A]\leq\int\lambda\left(\overline{\overline{a}},\eta\right)\left(\overline{\overline{a}}-a(\eta)\right)d\eta. \tag{27}$$

Proof. See Appendix. ■

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

Theorem 3 *(Maximum effort principle). Assume that* $\forall\eta,\forall a\leq\overline{\overline{a}},\ \partial_1 b(a,\eta)f(\eta)\geq\lambda\left(\overline{\overline{a}},\eta\right)$, *i.e. the marginal benefit of effort is sufficiently large. Then, the optimal plan is to implement maximum effort,* $A(\eta)=\overline{\overline{a}}$.

Proof. For any plan,

$$\begin{aligned}\int\left(b\left(\overline{\overline{a}},\eta\right)-b\left(a(\eta),\eta\right)\right)f(\eta)\,d\eta &\geq\int\inf_a\partial_1 b(a,\eta)\left(\overline{\overline{a}}-a(\eta)\right)f(\eta)\,d\eta\\ &\geq\int\lambda\left(\overline{\overline{a}},\eta\right)\left(\overline{\overline{a}}-a(\eta)\right)d\eta\\ &\geq C\left[\overline{\overline{A}}\right]-C[A]\end{aligned}$$

by Proposition 3. Hence,

$$\int b\left(\overline{\overline{a}},\eta\right)f(\eta)\,d\eta-C\left[\overline{\overline{A}}\right]\geq\int b\left(a(\eta),\eta\right)f(\eta)\,d\eta-C[A]$$

i.e., the principal's objective is maximized by inducing maximum effort. ■

Theorem 3 above shows that, if the marginal benefit of effort is sufficiently greater than the marginal cost, than maximum effort is optimal. A sufficient (although unnecessary) condition is for the firm to be sufficiently large. To demonstrate this, we parameterize the b function by $b(a,\eta)=Sb_*(a,\eta)$, where S is 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. Under this specification, the benefit of effort is multiplicative in firm size. This is plausible for most agent actions, which can be "rolled out" across the whole company and thus have a greater effect in a larger firm. Examples include the

[17] The proof shows that we can take $\lambda\left(\overline{\overline{a}},\eta\right)=\max\left(\partial C\left[\overline{\overline{A}}\right]/\partial a(\eta),0\right)$. We use partial derivatives such $\partial C[A]/\partial a(\eta)$. Their meaning is traditional and is as follows. Under weak conditions, $C[\cdot]$ is differentiable A, in the sense that there is a function $\xi(\eta)$ (unique up to sets of measure 0) such that, for any $\{B(\eta)\}$, $\lim_{h\to 0}\left(C[A+hB]-C[A]\right)/h=\int\xi(\eta)B(\eta)\,d\eta$. Then, we define $\partial C[A]/\partial a(\eta)=\xi(\eta)$.

choice of strategy, the launch of new projects, or increasing production efficiency.[18]

Let $\overline{F}$ denote the complementary cumulative distribution function of η, i.e. $\overline{F}(x) = \Pr(\eta \geq x)$. We assume that $\sup_\eta \overline{F}(\eta)/f(\eta) < \infty$ and $\inf_\eta \partial_1 b_*\left(\overline{\overline{a}}, \eta\right) > 0$, and define:

$$S_* = \frac{\Lambda\left(\overline{\overline{a}}\right)}{\inf_\eta \partial_1 b_*\left(\overline{\overline{a}}, \eta\right)}, \qquad \Lambda\left(\overline{\overline{a}}\right) \equiv \frac{g'(\overline{\overline{a}}) + g''(\overline{\overline{a}}) \sup_\eta \frac{\overline{F}(\eta)}{f(\eta)}}{v'\left(v^{-1}\left(u^{-1}(\underline{u}) + g(\overline{\overline{a}}) + (\overline{\eta} - \underline{\eta})g'(\overline{\overline{a}})\right)\right)}. \tag{28}$$

Calculations in the Online Appendix show that if, $S > S_*$, i.e. the firm is sufficiently large, then it is optimal for the principal to induce maximum effort. Indeed, in Proposition 3 we can take $\lambda\left(\overline{\overline{a}}, \eta\right) = \Lambda\left(\overline{\overline{a}}\right) f(\eta)$.

The intuition for the above is as follows. The numerator of $\Lambda\left(\overline{\overline{a}}\right)$ contains the two costs of inducing higher effort – the disutility imposed on the agent (the first term) plus the risk imposed by the incentive contract required to implement effort (the second term). These are scaled by the denominator, where the term in brackets is an upper bound on the pay received by the agent. The costs of effort are thus of similar order of magnitude to the agent's pay. The benefit of effort is enhanced firm value and thus of similar order of magnitude to firm size. If the firm is sufficiently large ($S > S_*$), the benefits of effort outweigh the costs and so maximum productive effort is optimal. A simple numerical example illustrates. Consider a firm with a \$10b market value and, to be conservative, assume that maximum effort increases firm value by only 1%. Then, maximum effort creates \$100m of value, which vastly outweighs the agent's salary. Even if it is necessary to double the agent's salary to compensate him for the costs of increased effort, this is swamped by the benefits.

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 $u^{-1}(\underline{u}) + g(\overline{\overline{a}}) + (\overline{\eta} - \underline{\eta})g'(\overline{\overline{a}})$, lowers α, and raises $\sup \overline{F}/f$. (For example, if the noise is uniform, then $\sup \overline{F}/f = \overline{\eta} - \underline{\eta}$.) Second, it is increasing in the agent's risk aversion parameterized by v and thus the risk imposed by incentives. 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 $g''(\overline{\overline{a}})$. Fourth, it is decreasing in the marginal benefit of effort $\left(\inf_\eta \partial_1 b_*\left(\overline{\overline{a}}, \eta\right)\right)$. Thus, the maximum effort principle is especially likely to hold if noise, risk aversion and the cost of effort are small.

We conjecture that a "maximum effort principle" holds under more general conditions than those considered above. For instance, it likely continues to hold if the principal's objective function is $\max_{\{a(\eta)\}} \int b\left(A(\eta), \eta\right) f(\eta)\, d\eta - C[A]$ 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

[18] Bennedsen, Perez-Gonzalez and Wolfenzon (2009) provide empirical evidence that CEOs have the same percentage effect on firm value, regardless of firm size; Edmans, Gabaix and Landier (2009) show that a multiplicative production function is necessary to generate empirically consistent predictions for the scaling of various measures of incentives with firm size.

slight variant is economically very similar, since the principal never wishes to implement $A(\eta) > \overline{\overline{a}}$ in our setting, but substantially more complicated mathematically, because the agent's action space now has boundaries and so the incentive constraints become inequalities. We leave this extension to future research. Hellwig (2007) shows that this reason alone is sufficient for a boundary effort level to be always optimal in a multiperiod discrete model and a continuous-time model that can be approximated by a discrete-time model, even in the absence of condition on the benefit of effort featured in this paper. Since the incentive constraints are inequalities with a boundary effort level, the principal has greater freedom in choosing the contract, which allows her to select a cheaper contract. Thus, the maximum effort result holds in settings even without a large benefit of effort. Lacker and Weinberg (1989) similarly derive a condition under which maximum effort (zero diversion in their setting) is optimal, for the case $v(c) = c$. In DeMarzo and Sannikov (2006), DeMarzo and Fishman (2007) and Biais et al. (2007), zero diversion is optimal since the agent is risk-neutral and so there is no trade-off between risk and incentives. Edmans, Gabaix, Sadzik and Sannikov (2009) extend the maximum effort principle to general T, for the case where $v(c) = \ln c$ (multiplicative preferences) and u is CRRA.

In the full contracting problem, which solves for both the optimal effort level and the cheapest implementing contract, tractable contracts are attained by forcing a constant incentive slope on the agent to rule out the ambiguous reward for performance. This is achieved in our paper through two key mechanisms. First, we achieve a constant marginal cost of effort by implementing a constant target action. This requires the removal of dynamics so that the action that the principal wishes to implement is independent of prior period outcomes. Previous papers remove dynamics via removing wealth effects, so that the cost of implementing a given action is constant. For example, HM assume CARA utility and a pecuniary cost of effort, so that wealth has no effect on the agent's risk aversion, and has an identical effect on the felicity from cash and cost of effort. DeMarzo and Sannikov (2006), DeMarzo and Fishman (2007) and Biais et al. (2007) assume risk-neutrality, so that risk aversion is independent of wealth (it is always zero) and the marginal utility of money is constant. The key insight of this paper is that we can remove dynamics without removing wealth effects, and thus without imposing constraints on the utility function or the cost of effort. Specifically, a constant target action need not require the cost of implementing the action to be constant – it only requires changes in these costs to be small compared to the benefits of effort. If the benefits of effort are sufficiently large (e.g. the firm is big), maximum effort remains optimal regardless of how the cost of implementing effort changes over time. Thus, our formulation allows for wealth effects to exist (and thus the utility function to be unrestricted), while at the same time removing dynamics and thus achieving tractability because such effects are small. The main limitation of our setup is that, in order to relax the HM assumptions, we require a restriction on the benefit of effort for Theorem 3 to hold.

Second, our timing assumption forces the constant marginal cost of effort (which is a consequence of the constant action) to equal the marginal felicity from cash state-by-state, and thus requires the reward for performance to be the same after every noise realization. In sum, the paper provides a set of sufficient conditions under which simple contracts can be obtained – actions following noise and a large benefit of effort – which is quite different than considered in prior literature. They may therefore hold in settings where the alternative assumptions are not satisfied and tractability was previously believed to be unattainable.

Appendix E considers other sufficient conditions required for Proposition 3 to hold, which do not assume the benefit of effort is multiplicative in firm size. That section also shows that we can derive the optimal $\{A(\eta)\}$ in certain cases even where the maximum effort principle does not apply.

3.3 Determinants of the Maximum Effort Level

The previous section assumed that the maximum productive effort level $\overline{\overline{a}}$ is exogenous. This section allows the principal to choose it endogenously according to the environment. We extend the contracting game to two stages. In the first stage, the principal chooses $\overline{\overline{a}}$. In practice, this may be achieved by physical investment, training the agent, or organizational design. For example, building a larger plant gives the agent greater scope to add value; training the agent or choosing an organizational structure that gives him greater responsibility and freedom have the same effect. Since physical investment, training and organizational design are costly to reverse, we model this decision as irreversible. In the second stage, the game studied in the core model is played out. In this stage, the action a may respond to the noise η, but the maximum productive effort $\overline{\overline{a}}$ has been fixed.

The principal's payoff is:

$$\int b\left(\min\left(A(\eta), \overline{\overline{a}}\right), \eta, \overline{\overline{a}}\right) d\eta - C[A] \tag{29}$$

where $b\left(a, \eta, \overline{\overline{a}}\right)$ is weakly increasing in a and decreasing in $\overline{\overline{a}}$. Higher flexibility $\overline{\overline{a}}$ is costly to the principal – for instance, we could have $b\left(a, \eta, \overline{\overline{a}}\right) = b(a, \eta) - H\left(\overline{\overline{a}}\right)$, where $H\left(\overline{\overline{a}}\right)$ is the cost of implementing flexibility level $\overline{\overline{a}}$.

Before we state the result formally, we summarize it. Under conditions described below, in the second stage, the principal will wish to implement the contract in Theorem 1 with $a^* = \overline{\overline{a}}$, i.e. the maximum effort principle applies. In the first stage, when choosing $\overline{\overline{a}}$, she will trade off the costs and benefits of a higher maximum effort. For instance, in the examples at the end of this section, $\overline{\overline{a}}$ is decreasing in the agent's disutility of effort and the noise dispersion. A trade-off exists in the first stage because the costs and benefits of flexibility are of similar order of magnitude. For example, increasing plant size has a continuous effect on firm value and involves a significant cost, which is also a function of firm size. However, it does not exist in the second stage because the costs of effort are a function of the agent's salary, and the benefits are discontinuous. Once the plant has been built, the agent must run it fully efficiently to prevent significant value loss – even small imperfections will cause large reductions in value and so the marginal benefit of effort is high (analogous to Kremer's (1993) O-ring theory). Thus, this enriched game features a simple optimal contract (since the target action in the second stage in constant), but one which also responds to the comparative statics of the environment. It may thus be a potentially useful way of modeling various economic problems, to achieve tractability while at the same time generating comparative statics.

To proceed more formally, consider the two following problems.

Problem 1: maximize over $\overline{\overline{a}}$ and all unrestricted contracts:

$$\max_{\overline{\overline{a}},\{a(\eta)\}} E\left[b\left(\min\left(A\left(\eta\right),\overline{\overline{a}}\right),\eta,\overline{\overline{a}}\right)\right]-C\left[A\right].$$

Problem 2: maximize over $\overline{\overline{a}}$ and use the contract in Theorem 1 which implements $\overline{\overline{a}}$:

$$\max_{\overline{\overline{a}}} E\left[b\left(\overline{\overline{a}},\eta,\overline{\overline{a}}\right)\right]-C\left[\overline{\overline{a}}\right].$$

where $C\left[\overline{\overline{a}}\right]$ is the expected cost of the contract implementing a constant action $\overline{\overline{a}}$.

Problem 2 optimizes over only a scalar $\overline{\overline{a}}$, while Problem 1 optimizes over a whole continuum of contracts, including those that do not implement maximum effort. However, under some simple conditions, Problem 2 is not restrictive – both problems have the same solution.

Proposition 4 *(Maximum effort in two-stage game). Let a^{**} denote the value of $\overline{\overline{a}}$ in a solution to Problem 1, and assume that $a^{**} > \underline{a}$ and that $\forall\eta$, $\inf_a \partial_1 b\left(a,\eta,a^{**}\right) f\left(\eta\right) \geq \lambda\left(a^{**},\eta\right)$. Then, the solution of Problem 1 is the same solution as Problem 2: that is, the solution of the problem that implements $A\left(\eta\right)=\overline{\overline{a}}$ is also the solution of the unrestricted contract.*

Proof. Immediate given Theorem 3. At a^{**}, the principal wants to implement maximum effort, i.e. $a\left(\eta\right)=a^{**}$ for all η. ■

At first glance, the condition in Proposition 4 may appear restrictive, since verifying it requires solving Problem 1. However, sufficient conditions are simply $\inf_a \partial_1 b\left(a,\eta,a^{**}\right) f\left(\eta\right) \geq \lambda\left(a^{**},\eta\right)$ for all a^{**} and η. The value λ can be calculated up to an integral, so bounds are reasonably straightforward to check in a given setting.

Illustrations We now illustrate the contract and comparative statics in three examples, for specific cases of u and v. We define $B\left(a\right)=E\left[b\left(a,\eta,a\right)\right]$, the principal's expected payoff given target effort a. The optimal contract gives $c=v^{-1}\left(g'\left(a\right)\eta+k\right)$ where k satisfies $E\left[u\left(g'\left(a\right)\eta-g\left(a\right)+k\right)\right]=\underline{u}$. Using previous notation, $k=K+g'\left(a\right)a$. The expected cost of the contract is $C\left[a\right]\equiv E\left[c\left(r\right)\right]=E\left[v^{-1}\left(g'\left(a\right)\eta+k\right)\right]$. It is straightforward to show that $C\left[a\right]$ increases in target effort a, the agent's reservation utility $\underline{u}$, and the dispersion of noise η; the proof relies on the dispersion techniques used in this paper.

The principal's problem is:

$$\max_{a} B(a) - C[a] \tag{30}$$

and the optimal contract is the contract described in Theorem 1 implementing a constant a. This is a simple problem to solve in many applied settings.

Example 1. Consider $u(x) = x$, $v(x) = x^{\gamma}$, $\gamma \in (0,1]$. We have $k = g(a) + \underline{u}$, and the contract is $c(r) = (g'(a)(r-a) + g(a) + \underline{u})^{1/\gamma}$. The expected cost is[19]

$$C[a] = E\left[(g'(a)\eta + g(a) + \underline{u})^{1/\gamma}\right].$$

$C[a]$ can be obtained in closed form for various specific cases. For example, $\gamma = 1/2$ yields $C[a] = g'(a)^2 \sigma^2 + (g(a) + \underline{u})^2$; $\underline{u} = 0$ and $g(a) = e^{Ga}$ yields $C[a] = e^{Ga/\gamma} E\left[(G\eta + 1)^{1/\gamma}\right]$. The principal chooses a to maximize $B(a) - e^{Ga/\gamma} E\left[(G\eta + 1)^{1/\gamma}\right]$. Simple calculations show that the target action is decreasing in the marginal cost of effort G, risk aversion γ and the dispersion of noise η.

Example 2. Consider $v(x) = \ln x$ and $u(x) = e^{(1-\gamma)x}/(1-\gamma)$ for $\gamma > 1$, so that the utility function is $\left(ce^{-g(a)}\right)^{1-\gamma}/(1-\gamma)$, as is commonly used in macroeconomics: it is CRRA and multiplicative in consumption and effort. We also assume $\eta \sim N(0, \sigma^2)$. Then, the contract is $c(r) = \exp(g'(a)(r-a) + k)$ with $k = \ln \underline{c} + g(a) - (1-\gamma) g'(a)^2 \sigma^2/2$, where $u(\ln \underline{c})$ is the reservation utility. The expected cost of the contract is:

$$C[a] = \underline{c} \exp\left(g(a) + \gamma g'(a)^2 \sigma^2/2\right).$$

Again, calculations show that a is decreasing in the cost of effort, risk aversion and noise dispersion. We thus obtain the standard comparative statics, but for a contract that is log-linear, rather than linear in returns. Murphy (1999) argues that log-linear contracts are empirically more relevant.

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

Note that HM not only have a constant target action, but an additive effect of effort. We can obtain this result with $b\left(a, \eta, \overline{\overline{a}}\right) = a + \left(a - \overline{\overline{a}}\right) \beta\left(\overline{\overline{a}}, \eta\right)$, for some function $\beta\left(\overline{\overline{a}}, \eta\right) \geq \lambda\left(\overline{\overline{a}}, \eta\right)/f(\eta)$. In the second stage of the game, having chosen $\overline{\overline{a}}$, the principal wishes to implement constant effort $\overline{\overline{a}}$ for all η, because the marginal cost of shirking (parameterized by β) is sufficiently high. Moving to the first stage, since the principal knows that $a = \overline{\overline{a}}$ in the second stage, her benefit function is $b\left(a, \eta, \overline{\overline{a}}\right) = a$: effort has an additive effect.

The key complication in obtaining the HM result is reconciling the linear marginal benefit

[19] A variant is the case $u(x) = x$ and $v(x) = \ln x$. Then, the contract is $\ln c(r) = g'(a)(r-a) + g(a) + \underline{u}$, and the expected cost is $C(a) = \exp[g(a) + \underline{u}] E\left[e^{g'(a)\eta}\right]$.

of effort required for an additive effect, with the high marginal benefit of effort required for the maximum effort principle to apply to guarantee a constant action. The two-stage game resolves this tension because the marginal benefit of effort is moderate in the first stage and very high in the second stage, as discussed in the plant example earlier.

Under this formulation, the cost of the contract implementing $a = \overline{\overline{a}}$ is $C[a] = \underline{c} + \frac{1}{2}Ga^2 + \frac{\gamma}{2}G^2a^2\sigma^2$ and the principal maximizes $a - \underline{c} - \frac{1}{2}G\overline{\overline{a}}^2 - \frac{\gamma}{2}G^2a^2\sigma^2$ which yields the result $a = 1/G(1 + G\gamma\sigma^2)$, exactly as in HM. Thus, using the HM conditions of exponential utility, a pecuniary quadratic cost of effort and Gaussian noise in the above specification, leads to the same optimal contract (not just the implementation contract) as in HM.

In Appendix E, we also provide explicit conditions under which maximum effort is optimal for the three above examples, i.e. a specialization of the conditions in Proposition 4 to these cases. These conditions allow straightforward verification of whether the maximum effort principle holds.

4 Conclusion

This paper has identified and analyzed a class of multiperiod situations in which the optimal contract is tractable, without requiring exponential utility, a pecuniary cost of effort, Gaussian noise or continuous time. The contract's functional form is independent of the agent's utility function, reservation utility and noise distribution. Furthermore, when the cost of effort can be expressed in financial terms, the optimal contract is linear and so the slope, in addition to the functional form, is independent of these parameters.

The key to tractability in discrete time is specifying the noise before the action in each period, which forces the incentive compatibility constraint to hold state-by-state rather than just on average, and tightly constraints the set of contracts available to the principle. The optimal contract is very similar in continuous time, where noise and actions occur simultaneously. Hence, if underlying reality is continuous time, it is best approximated in discrete time under our timing assumption. Moving to the full contracting problem, our two-stage model allows the principal to choose the target effort level to respond to the details of the environment, while retaining tractability. The principal initially sets a lower maximum productive effort level if the agent is more risk averse or faces a higher cost of effort or greater noise. However, in each subsequent period, the principal wishes the agent to exert maximum effort, regardless of how output evolves. If the benefits of effort are sufficiently high (e.g. the firm is much larger than the agent's salary), they swamp the costs, and so the optimal effort level is independent of how the agent's wealth evolves over time.

Our paper suggests several avenues for future research. The HM framework has proven valuable in many areas of applied contract theory owing to its tractability; however, some models have used the HM result in settings where the assumptions are not satisfied (see the critique of Hemmer (2004)). Our framework allows tractable contracts to be achieved in such situations. 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), effort is non-pecuniary, or noise is not Gaussian (e.g. is bounded). While we considered the specific application of executive compensation, other possibilities include bank regulation, team production, insurance or taxation.[20] In ongoing work (Edmans, Gabaix, Sadzik and Sannikov (2009)) we extend tractable contracts to a dynamic setting where the agent consumes in each period, can privately save, and may smooth earnings intertemporally.

In addition, while our model has relaxed a number of assumptions required for tractability, it continues to impose a number of restrictions. In particular, the optimal action can only be solved tractably if the maximum effort principle applies or in certain other cases (e.g. linear cost of effort). Grossman and Hart (1983) and Garrett and Pavan (2009) show that solving for the optimal action in a general case is typically extremely complex; whether we can extend tractability to broader settings is an important area for future research. Similarly, while Section 3 allows for the action to depend on the noise in period t, a useful extension would be to allow the action to depend on the full history of outcomes. Other restrictions are mostly technical rather than economic. For example, our multiperiod model assumes independent noises with log-concave density functions; and our extension to noise-dependent target actions assumes an open action set where the maximum feasible effort level exceeds the maximum productive effort level. Some of these assumptions may not be valid in certain situations, limiting the applicability of our framework. Further research may be able to broaden the current setup.

[20] See Golosov, Kocherlakota and Tsyvinski (2003) and Farhi and Werning (2009) for taxation applicaitons of the principal-agent problem.

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, defined over r or η
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
$C[A]$	Expected cost of contract implementing $\left\{A(\eta), \eta \in \left(\underline{\eta}, \overline{\eta}\right)\right\}$
$\overline{F}$	Complementary cumulative distribution function of η
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, defined over r or η

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 IC constraints bind. Shaked and Shanthikumar (2007, Section 3.B) provide an excellent summary of known facts about this concept. This section provides a self-contained guide of the relevant results for our paper, as well as proving some new results.

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

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

Lemma 1 *Let X be a random variable and f, h be functions such that $0 \leq f(y) - f(x) \leq h(y) - h(x)$ whenever $x \leq y$. Then $f(X)$ is less dispersed than $h(X)$.*

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

Lemma 2 *Assume that X is less dispersed than Y and let f denote a weakly increasing function, h a weakly increasing concave function, and ϕ a weakly increasing convex function. Then:*

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

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

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

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

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

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

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

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

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

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:*

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

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

We use the following Lemma to allow us to integrate inequalities with subderivatives. All the Lemmas in this subsection are proven in the Online Appendix.

Lemma 4 *Assume that, over an interval I: (i) $f'_-\left(x\right)\geq j\left(x\right)\ \forall\ x$, for an continuous function $j\left(x\right)$ and (ii) there is a C^1 function h such that $f+h$ is nondecreasing. Then, for two points $a\leq b$ in I, $f\left(b\right)-f\left(a\right)\geq \int_a^b j\left(x\right)dx$.*

Condition (ii) prevents $f\left(x\right)$ from exhibiting discontinuous downwards jumps, which would prevent integration.[21]

The following Lemma is the chain rule for subderivatives.

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 $f\left(x\right)$, with $h'\left(f\left(x\right)\right)>0$. Then, $\left(h\circ f\right)'_-\left(x\right)=h'\left(f\left(x\right)\right)f'_-\left(x\right)$.*

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

Lemma 6 *Let x be a real number and f, h be functions defined in a left neighborhood of x. Then $\left(f+h\right)'_-\left(x\right)\geq f'_-\left(x\right)+h'_-\left(x\right)$. When h is differentiable at x, then $\left(f+h\right)'_-\left(x\right)=f'_-\left(x\right)+h'\left(x\right)$.*

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

B Detailed Proofs

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

Proof of Theorem 1

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

1). Redundancy of Messages

Let $\mathbf{r}$ denote the vector $(r_1, ..., r_T)$ and define $\boldsymbol{\eta}$ and $\mathbf{a}$ analogously. Define $\mathbf{g}(\mathbf{a}) = g(a_1) + ... + g(a_T)$. Let $\widetilde{V}_M(\mathbf{r}, \boldsymbol{\eta}) = v(\widetilde{c}(\mathbf{r}, \boldsymbol{\eta}))$ denote the felicity given by a message-dependent contract if the agent reports $\boldsymbol{\eta}$ and the realized signals are $\mathbf{r}$. Under the revelation principle, we can restrict the analysis to mechanisms that induce the agent to truthfully report the noise $\boldsymbol{\eta}$. The incentive compatibility (IC) constraint is that the agent exerts effort $\mathbf{a}$ and reports $\widehat{\boldsymbol{\eta}} = \boldsymbol{\eta}$:

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

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

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

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:

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

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

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

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

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

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

We will prove the Theorem by induction on T.

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

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

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

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

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

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

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

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

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 $g'\left(a^{*}\right)$. Therefore, the left derivative $V'_{left}\left(r\right)$ exists, and equals $g'\left(a^{*}\right)$. Second, consider $r < r'$. Dividing (36) through by $r - r'$, and taking the limit $r' \downarrow r$ shows that the right derivative $V'_{right}\left(r\right)$ exists, and equals $g'\left(a^{*}\right)$. Therefore,

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

Since r has interval support[22], we can integrate to obtain, for some integration constant K:

$$V(r) = g'(a^*)\, r + K. \tag{39}$$

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

Note that the above proof (for deterministic contracts where $a_t^* < \overline{a}$) does not require log-concavity of η_t, nor that u satisfies NIARA. This is because the contract (7) is the only incentive compatible contract. These assumptions are only required for the general proof, where other contracts (e.g. randomized ones) are also incentive compatible, to show that they are costlier than contract (7).

3). General Proof

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

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

$$\frac{d}{d\varepsilon_-}\mathrm{E}\left[u\left(\widetilde{V}(\widetilde{r}_1, ..., \widetilde{r}_t + \varepsilon, ..., \widetilde{r}_T)\right) \mid \widetilde{r}_1, .., \widetilde{r}_t\right]_{|\varepsilon=0} \geq g'(a_t^*)\,\mathrm{E}\left[u'\left(\widetilde{V}(\widetilde{r}_1, ..., \widetilde{r}_t, ..., \widetilde{r}_T)\right) \mid \widetilde{r}_1, .., \widetilde{r}_t\right] \tag{40}$$

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 are costlier. 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 $a_t^* = \overline{a}$.

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

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

$$V'_-(r) \geq g'(a^*). \tag{41}$$

[22] 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.

It is intuitive that (41) should bind, as this minimizes the variability in the agent's pay and thus constitutes efficient risk-sharing. We now prove that this is indeed the case; to simplify exposition, we normalize $g(a^*) = 0$ w.l.o.g.[23] If constraint (41) binds, the contract is $V^0(r) = g'(a^*)\,r + K$, where K satisfies $\mathrm{E}\left[u\left(g'(a^*)\,r + K\right)\right] = \underline{u}$. We wish to show that any other contract $V(r)$ that satisfies (41) is weaklier costlier.

By assumption (iii) in Statement Σ_1, V is nondecreasing. We can therefore apply Lemma 4 to equation (41), where condition (ii) of the Lemma is satisfied by $h(r) \equiv 0$. This implies that for $r \leq r'$, $V(r') - V(r) \geq g'(a^*)(r' - r) = V^0(r') - V^0(r)$. Thus, using Lemma 1, $V(\widetilde{r})$ is more dispersed than $V^0(\widetilde{r})$.

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

$$\mathrm{E}\left[u\left(V\left(\widetilde{r}\right)\right)\right] \geq \underline{u} = \mathrm{E}\left[u\left(V^0\left(\widetilde{r}\right)\right)\right]. \tag{42}$$

Applying Lemma 2 to the convex function $\phi \circ u^{-1}$ and inequality (42), 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 item 2 in Section 2.3, the extension to a risk-averse principal. Suppose that the principal wants to minimize $\mathrm{E}\left[w(c)\right]$, where w is an increasing and concave function, rather than $\mathrm{E}[c]$. Then, the above contract is optimal if $w \circ v^{-1} \circ u^{-1}$ is convex, i.e. $u \circ v \circ w^{-1}$ is concave. This requires w to be "not too concave," i.e. the agent to be not too risk-averse.

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

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

$$\widehat{u}(x) = \mathrm{E}\left[u\left(x + g'\left(a^*_{T+1}\right)\widetilde{r}_{T+1}\right)\right]. \tag{43}$$

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 (44) below:

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

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

k represents the expected felicity from contract V based on all noise realizations up to and including time T.

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

Taking expectations of (44) over $\widetilde{r}_1, ..., \widetilde{r}_T$ yields:

$$\mathrm{E}\left[\widehat{u}\left(k\left(\widetilde{r}_1, ..., \widetilde{r}_T\right)\right)\right] = \mathrm{E}\left[u\left(V\left(\widetilde{r}_1, ..., \widetilde{r}_{T+1}\right)\right)\right] \geq \underline{u}, \tag{45}$$

where the inequality comes from Condition (i) for utility function u and contract V at time $T+1$. Hence, Condition (i) holds for utility function $\widehat{u}$ and contract k at time t. In addition, it is immediate that $\mathrm{E}\left[\widehat{u}\left(k\left(\widetilde{r}_1, ..., \widetilde{r}_T\right)\right) \mid \widetilde{r}_1, .., \widetilde{r}_t\right]$ is nondecreasing in $\widetilde{r}_t$. (Condition (iii)). We thus need to show that Condition (ii) is satisfied.

Since equation (40) 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:

$$\frac{dV}{dr_{T+1\,-}}\left(r_1, ..., r_{T+1}\right) \geq g'\left(a^*_{T+1}\right). \tag{46}$$

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

Using (43), we can rewrite equation (44) as

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

Since u exhibits NIARA, $-u''(x)/u'(x)$ is nonincreasing in x. This is equivalent to $u' \circ u^{-1}$ being weakly convex. We can thus apply Lemma 2 to yield:

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

$$\mathrm{E}_T\left[u'\left(V\left(\widetilde{r}_1, ..., \widetilde{r}_{T+1}\right)\right)\right] \geq \mathrm{E}_T\left[\widehat{u}'\left(k\left(\widetilde{r}_1, ..., \widetilde{r}_T\right)\right)\right]. \tag{47}$$

Applying definition (44) to the left-hand side of Condition (ii) for $T+1$ yields, with $t = 1...T$,

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

Taking expectations of equation (47) at time t and substituting into the right-hand side of the above equation yields:

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

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

$$\widehat{\phi}\left(x\right)\equiv\mathrm{E}\left[\phi\left(x+g'\left(a_{T+1}\right)\widetilde{r}_{T+1}\right)\right]. \tag{48}$$

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

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

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

$$C_k=\mathrm{E}\left[\widehat{\phi}\left(k\left(\widetilde{r}_1,...,\widetilde{r}_T\right)\right)\right]\geq C_{V^0}=\mathrm{E}\left[\phi\left(\sum_{t=1}^{T+1}g'\left(a_t^*\right)\widetilde{r}_t+K\right)\right]. \tag{49}$$

Using equation (48) yields:

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

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 (44) 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

$$\mathrm{E}_t\left[\phi\left(V\left(\widetilde{r}_1,...,\widetilde{r}_{T+1}\right)\right)\right]\geq\mathrm{E}_t\left[\phi\left(k\left(\widetilde{r}_1,...,\widetilde{r}_T\right)+g'\left(a_{T+1}^*\right)\widetilde{r}_{T+1}\right)\right]$$
$$\mathrm{E}\left[\phi\left(V\left(\widetilde{r}_1,...,\widetilde{r}_{T+1}\right)\right)\right]\geq\mathrm{E}\left[\phi\left(k\left(\widetilde{r}_1,...,\widetilde{r}_T\right)+g'\left(a_{T+1}^*\right)\widetilde{r}_{T+1}\right)\right]=C_k\geq C_{V^0}.$$

where the final inequality comes from (49). 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:

$$u\left(\overline{V}\left(r_1,...,r_T\right)\right)\equiv\mathrm{E}_T\left[u\left(\widetilde{V}\left(r_1,...,r_T\right)\right)\right]. \tag{50}$$

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 that Condition (iii) holds for $\overline{V}$, given it holds for $\widetilde{V}$. We thus need to show that Condition (ii) is also satisfied. Applying Jensen's inequality to equation (50) and the function $u' \circ u^{-1}$ (which is convex since u exhibits NIARA) 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 take expectations to obtain

$$\mathrm{E}_t\left[u'\left(\widetilde{V}\left(\widetilde{r}_1, ..., \widetilde{r}_T\right)\right)\right] \geq \mathrm{E}_t\left[u'\left(\overline{V}\left(\widetilde{r}_1, ..., \widetilde{r}_T\right)\right)\right]. \tag{51}$$

Applying definition (50) to the left-hand side of (40) 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 (51) 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 (50) 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

$$\overline{u}\left(x\right) \equiv u\left(x - \sum_{s=1}^{T} g\left(a_s^*\right)\right). \tag{52}$$

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.

$$0 \in \arg\max_{\varepsilon} \mathrm{E}_t\, u\left(\widetilde{V}\left(\widetilde{r}_1, ..., \widetilde{r}_t + \varepsilon, ..., \widetilde{r}_T\right) - g\left(a_t^* + \varepsilon\right) - \sum_{s=1...T, s\neq t} g\left(a_s^*\right)\right). \tag{53}$$

We note that, for a function $f(\varepsilon)$, $0 \in \arg\max_\varepsilon f(\varepsilon)$ implies that for all $\varepsilon < 0$, $\left(f(0) - f(\varepsilon)\right)/(-\varepsilon) \geq 0$, hence, taking the $\liminf_{y\uparrow 0}$, we obtain $\frac{d}{d\varepsilon_-} f'_-\left(\varepsilon\right)_{|\varepsilon=0} \geq 0$. Call $X(\varepsilon)$ the argument of u in equation (53). Applying this result to (53), 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 $\mathrm{E}_t\left[u'\left(X(0)\right)\left(\frac{d}{d\varepsilon_-}X\left(\varepsilon\right)_{|\varepsilon=0}\right)\right] \geq 0$. Using Lemma 6, $\frac{d}{d\varepsilon_-}X\left(\varepsilon\right)_{|\varepsilon=0} = \frac{d}{d\varepsilon_-}\widetilde{V}\left(\widetilde{r}_1, ..., \widetilde{r}_t + \varepsilon, ..., \widetilde{r}_T\right) - g'\left(a_t^*\right)$, hence we obtain:

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

Using again Lemma 5, this can be rewritten:

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

i.e., using the notation (52),

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

Therefore, Condition (ii) of Statement Σ_T holds.

Finally, we verify Condition (iii). Apply (53) to signal r_t and deviation $\varepsilon < 0$. We obtain:

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

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

We can now apply Statement Σ_T to contract $\widetilde{V}$ and function $\overline{u}$, to prove that any globally IC contract is weakly costlier than contract $V^0 = \sum_{t=1}^{T} g'(a_t^*) r_t + K$. Moreover, it is clear that V^0 satisfies the global IC conditions in equation (53). Thus, V^0 is the cheapest contract that satisfies the global IC constraint.

Proof of Proposition 1

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

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

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

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

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

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

This can be rewritten

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

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

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

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

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

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

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

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

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

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

Proof of Theorem 2

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

Lemma 7 *Consider a standard Brownian process Z_t with filtration $\mathcal{F}_t$, a deterministic non-negative process α_t, an $\mathcal{F}_t-$adapted process β_t, $T \geq 0$, $X = \int_0^T \alpha_t dZ_t$, 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.

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

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

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

We proceed in two steps.

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

$$a_t^* \in \arg\max_{a_t} \theta_t a_t - g(a_t). \tag{54}$$

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

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

Suppose that (54) is not verified on the set τ of times with positive measure. Then, consider a policy a such that $\theta_t a_t - g(a_t) > \theta_t a_t^* - g(a_t^*)$ for $t \in \tau$, and $a_t = a_t^*$ on $[0,T] \setminus \tau$. We thus have $L > 0$. Consider the agent's utility under policy a:

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

where U^{a*} is the agent's utility under policy a^*. Hence, as $U^a > U^{a*}$, the IC condition is violated. We conclude that condition (54) is necessary for the contract to satisfy the IC condition.

We next show that condition (54) is also sufficient to satisfy the IC condition. Indeed, consider any adapted policy a. Then, $L \leq 0$. So, the above reasoning shows that $U^a \leq U^{a^*}$. Policy a^* is at least as good as any alternative strategy a.

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

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

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

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

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

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

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

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

We wish to show that a contract $v_T = Y + K_Y$, with $\mathrm{E}[u(Y + K_Y)] \geq \underline{u}$, has a weakly greater expected cost than a contract $v = X + K_X$, with $\mathrm{E}[u(X + K_X)] = \underline{u}$. Lemma 7 implies that $\mathrm{E}[u(X + K_X)] \geq \mathrm{E}[u(Y + K_X)]$, and so

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

Thus, $K_X \leq K_Y$. Since v is increasing and concave, v^{-1} is convex and $-v^{-1}$ is concave. We can therefore apply Lemma 7 to function $-v^{-1}$ to yield:

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

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

Proof of Proposition 2

The proof is by induction.

Proof of Proposition 2 for $T = 1$. We remove time subscripts and let $V(\widehat{\eta}) = v(C(\widehat{\eta}))$ 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: $\forall \eta, \forall \widehat{\eta}$,

$$V(\widehat{\eta}) - g(A(\widehat{\eta}) + \widehat{\eta} - \eta) \leq V(\eta) - g(A(\eta)). \tag{55}$$

Defining

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

the truth-telling constraint (55) can be rewritten,

$$g(A(\widehat{\eta})) - g(A(\widehat{\eta}) + \widehat{\eta} - \eta) \leq \psi(\eta) - \psi(\widehat{\eta}). \tag{56}$$

Rewriting this inequality interchanging η and $\widehat{\eta}$ and combining with the original inequality (56) yields:

$$\forall \eta, \forall \widehat{\eta}: g(A(\widehat{\eta})) - g(A(\widehat{\eta}) + \widehat{\eta} - \eta) \leq \psi(\eta) - \psi(\widehat{\eta}) \leq g(A(\eta) + \eta - \widehat{\eta}) - g(A(\eta)). \tag{57}$$

Consider a point η where A is continuous and take $\widehat{\eta} < \eta$. Dividing (57) 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 (57) by $\eta - \widehat{\eta} < 0$ and taking the limit $\widehat{\eta} \downarrow \eta$ yields $\psi'_{right}(\eta) = g'(A(\eta))$. Hence,

$$\psi'(\eta) = g'(A(\eta)), \tag{58}$$

at all points η where A is continuous.

Equation (58) holds only almost everywhere, since we have only assumed that A is almost everywhere continuous. To complete the proof, we require a regularity argument about ψ (otherwise ψ might jump, for instance). We will show that ψ is absolutely continuous (see, e.g., Rudin (1987), p.145). Consider a compact subinterval I, and $\overline{a}_I = \sup\{A(\eta) + \eta - \widehat{\eta} \mid \eta, \widehat{\eta} \in I\}$, which is finite because A is assumed to be bounded in any compact subinterval of η. Then, equation (57) implies:

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

This implies that ψ is absolutely continuous on I. Therefore, by the fundamental theorem of calculus for almost everywhere differentiable functions (Rudin (1987), p.148), we have that for any η, η_*, $\psi(\eta) = \psi(\eta_*) + \int_{\eta_*}^{\eta} \psi'(x)\,dx$. From (58), $\psi(\eta) = \psi(\eta_*) + \int_{\eta_*}^{\eta} g'(A(x))\,dx$, i.e.

$$V(\eta) = g(A(\eta)) + \int_{\eta_*}^{\eta} g'(A(x))\,dx + k \tag{59}$$

with $k = \psi(\eta_*)$. This concludes the derivation of the contract when $T = 1$.

"*Second-order conditions.*" We next show that the contract (59) does implement effort $A(\eta)$, iff $A(\eta) + \eta$ is nondecreasing: we have verified the first order condition, but we need to show that (55) holds given the proposed contract, that is, that $\Phi(\widehat{\eta}) \equiv V(\widehat{\eta}) - g(A(\widehat{\eta}) + \widehat{\eta} - \eta)$ has a maximum at η.

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

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

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

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

Proof that $A(\eta) + \eta$ nondecreasing is a necessary condition. Call $R(\eta) = A(\eta) + \eta$. Suppose by contradiction that there are two points $\eta < \eta'$ such that $R(\eta) > R(\eta')$. Those two points can be taken arbitrarily close (indeed, consider a large N, the points $\eta_i = \eta + (\eta' - \eta)\,i/N$, $i = 0...N$; there must be an i such that $R(\eta_i) > R(\eta_{i+1})$, otherwise we would have $R(\eta) = R(\eta_0) \leq R(\eta_N) = R(\eta')$). As domain $\mathcal{A}$ of actions is open, that implies that $A(\eta) + \eta - \eta' \in \mathcal{A}$. Applying (55) at point η and η', we have:

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

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

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

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

$$\eta_{T+1}\in\arg\max_{\widehat{\eta}_{T+1}}V\left(\eta_1,...,\eta_T,\widehat{\eta}_{T+1}\right)-g\left(A\left(\widehat{\eta}_{T+1}\right)+\widehat{\eta}_{T+1}-\eta_{T+1}\right)-\sum_{t=1}^{T}g\left(a_t^*\right). \tag{60}$$

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

$$V\left(\eta_1,...,\eta_T,\widehat{\eta}_{T+1}\right)=W_{T+1}\left(\widehat{\eta}_{T+1}\right)+k\left(\eta_1,...,\eta_T\right), \tag{61}$$

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

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

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

Defining

$$\widehat{u}\left(x\right)=\mathrm{E}\left[u\left(x+W_{T+1}\left(\widetilde{\eta}_{T+1}\right)\right)\right], \tag{62}$$

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

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

for some constant k_*. Combining this with (59), the only incentive compatible contract is:

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

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

Proof of Proposition 3

Step 1. It is easier to work in terms of $Q\left(\eta\right)=g'\left(A\left(\eta\right)\right)$, the marginal cost of effort associated with plan $A(\eta)$. With a slight abuse of notation, define $C[Q]$ as the expected cost of implementing plan $Q = \{Q(\eta)\}$. From Proposition 2 with $T = 1$, $c(\eta, Q) = v^{-1}\left(g \circ (g')^{-1}(Q(\eta)) + \int_0^\eta Q(x)\,dx + K\right)$, where K is the solution of $E\left[u\left(\int_0^\eta Q(x)\,dx + K\right)\right] = \underline{u}$. Then, the expected cost is: $C[Q] = E[c(\eta, Q)]$.

We first establish that the contract cost $C[Q]$ is convex in the plan Q. Consider two plans Q^1 and Q^2, $\theta_1 + \theta_2 = 1$ with $\theta_1, \theta_2 \in [0,1]$, and the plan Q defined by $Q(\eta) = \theta_1 Q_1(\eta) + \theta_2 Q_2(\eta)$. Since u is concave,

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

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

Step 2. Since C is convex, we have:

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

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

C 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 $v(c) - \left[g^F(a) + G(m)\right]$, where g, G are increasing and convex, $G(0) = 0$, and $G'(0) \geq g'\left(\overline{\overline{a}}\right)$. The final assumption means that manipulation is costlier than fundamental effort.

We define $a = a^F + m$ and the cost function $g(a) = \min_{a^F, M}\left\{g^F(a) + G(m) \mid a^F + m = a\right\}$, so that $g(a) = g^F(a)$ for $a \in (\underline{a}, \overline{\overline{a}}]$ and $g(a) = g^F(a) + g(m - a)$ for $a \geq \overline{\overline{a}}$, which is increasing and convex. Then, firm value can be written $b\left(\min\left(a, \overline{\overline{a}}\right), \widetilde{\eta}\right)$, as in equation (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 Proposition 2 implements $a \leq \overline{\overline{a}}$, so the agent will not engage in manipulation. Therefore, the rational expectations price is $P_1 = e^r$.

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

References

[1] Arnott, R. and J. Stiglitz (1988): "Randomization with Asymmetric Information." *RAND Journal of Economics* 19, 344-362

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

[3] Bennedsen, M., F. Perez-Gonzalez and D. Wolfenzon (2009): "Do CEOs Matter?" Working Paper, Copenhagen Business School

[4] Biais, B., T. Mariotti, G. Plantin and J.-C. Rochet (2007): "Dynamic Security Design: Convergence to Continuous Time and Asset Pricing Implications." *Review of Economic Studies* 74, 345-390

[5] Biais, B., T. Mariotti, J.-C. Rochet and S. Villeneuve (2009): "Large Risks, Limited Liability and Dynamic Moral Hazard." Working Paper, Université de Toulouse

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

[7] Cooley, T. and E. Prescott (2005): "Economic Growth and Business Cycles," in "Frontiers in Business Cycle Research," T. Cooley ed., Princeton University Press, Princeton

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

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

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

[11] Dittmann, I., E. Maug and O. Spalt (2009): "Sticks or Carrots? Optimal CEO Compensation when Managers are Loss-Averse." *Journal of Finance*, forthcoming

[12] Edmans, A., X. Gabaix and A. Landier (2009): "A Multiplicative Model of Optimal CEO Incentives in Market Equilibrium." *Review of Financial Studies*, forthcoming

[13] Edmans, A., X. Gabaix, T. Sadzik and Y. Sannikov (2009): "Dynamic Incentive Accounts." NBER Working Paper No. 15324.

[14] Farhi, E. and I. Werning (2009): "Capital Taxation: Quantitative Explorations of the Inverse Euler Equation." Working Paper, Harvard University.

[15] Golosov, M., N. Kocherlakota and A. Tsyvinski (2003): "Optimal Indirect Capital Taxation." *Review of Economic Studies*, 70, 569–587.

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

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

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

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

[20] He, Z. (2009a): "Optimal Executive Compensation when Firm Size Follows Geometric Brownian Motion." *Review of Financial Studies* 22, 859-892.

[21] He, Z. (2009b): "Dynamic Compensation Contracts with Private Savings." Working Paper, University of Chicago

[22] Hellwig, M. (2007): "The Role of Boundary Solutions in Principal-Agent Problems of the Holmstrom-Milgrom Type." *Journal of Economic Theory* 136, 446-475

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

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

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

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

[27] Karatzas, I. and S. E. Shreve (1991): *Brownian Motion and Stochastic Calculus*, 2nd edition, Springer Verlag

[28] Kremer, m. (1993): "The O-Ring Theory of Economic Development." *Quarterly Journal of Economics* 108, 551-576.

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

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

[31] Laffont, J.-J. and D. Martimort (2002): "The Theory of Incentives: The Principal-Agent Model." Princeton University Press, Princeton.

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

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

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

[35] Murphy, K. (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

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

[37] Phelan, C. and R. Townsend (1991): "Private Information and Aggregate Behaviour: Computing Multi-Period, Information-Constrained Optima." *Review of Economic Studies* 58, 853-881

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

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

[40] Rudin, W. (1987): *Real and Complex Analysis*, 3rd edition, McGraw-Hill

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

[42] Sappington, D. (1983): "Limited Liability Contracts Between Principal and Agent." *Journal of Economic Theory* 29, 1-21

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

[44] Shaked, M. and G. Shanthikumar (2007): *Stochastic Orders*, Springer Verlag

[45] Spear, S. and S. Srivastava (1987): "On Repeated Moral Hazard With Discounting." *Review of Economic Studies* 54, 599-617

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

Online Appendix for "Tractability in Incentive Contracting"

Alex Edmans and Xavier Gabaix

November 9, 2009

D Multidimensional Signal and Action

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

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

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

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

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

i.e., explicitly:

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

The following contract is optimal. The agent is paid

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

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

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

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

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

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

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

and so:

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

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

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

We now analyze two specific applications of this extension.

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

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

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

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

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

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

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

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

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

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

E Extension to The Optimal Effort Level

E.1 Illustrations for Proposition 2

E.1.1 Affine Cost of Effort

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

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

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

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

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

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

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

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

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

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

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

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

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

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

E.1.2 Exponential u and Linear v

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

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

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

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

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

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

so the principal solves a collection of problems

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

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

for (possibly) varying $\underline{u}(\eta_1, \ldots, \eta_{T-1})$. By concavity, the solutions of these problems for each $\underline{u}(\eta_1, \ldots, \eta_{T-1})$ are unique. Moreover, this uniqueness implies that the solutions for different values of $\underline{u}(\eta_1, \ldots, \eta_{T-1})$ may differ only in the constant K. Therefore, the optimal target action $A_T(\eta_1, \ldots, \eta_{T-1}, \eta_T)$ does not depend on $\eta_1, \ldots, \eta_{T-1}$.

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

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

$$\max_{a(\cdot),y(\cdot)} \int_{\underline{\eta}}^{\overline{\eta}} \left(Ba(x) - \frac{1}{2}Ga(x)^2 - y(x)\right) dx$$

$$\text{s.t.} \quad \int_{\underline{\eta}}^{\overline{\eta}} \left(-e^{-\gamma y(x)}\right) = \underline{u},$$

$$y'(x) = a(x).$$

The Lagrangian of this problem is

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

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

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

Substituting the first equality into the second we get

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

Rearranging and taking a logarithm gives

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

Differentiating the last equality gives

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

which can be simplified into

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

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

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

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

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

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

and the first-order conditions are

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

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

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

we obtain

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

Rearranging and taking a logarithm gives

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

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

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

E.2 Conditions for Maximum Effort Principle

Section 3.2 showed that the condition in Theorem 3,

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

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

By Proposition 2, the optimal contract is:

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

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

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

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

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

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

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

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

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

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

which implies (71).

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

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

and so the additional pay is:

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

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

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

Second, the upper bound for $\frac{\partial C[A]}{\partial a(\eta)}$ and thus $\lambda\left(\overline{\overline{a}},\eta\right)$ is simpler when noise is bounded both above and below. If $supp\,\eta=[\underline{\eta},\overline{\eta}]$ and $g'''(x)\geq0$ for all x. Then

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

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

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

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

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

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

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

Let $\overline{c}=v^{-1}\left(u^{-1}(\underline{u})+g(\overline{\overline{a}})+(\overline{\eta}-\underline{\eta})g'(\overline{\overline{a}})\right)$. Then, all on the equilibrium consumptions are no greater than $\bar{c}$. Hence, the terms in inequality (71) can be bounded as

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

which gives the claimed inequality.

E.3 Illustrations for Proposition 4

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

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

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

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

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

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

Inequality (72) establishes the bound

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

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

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

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

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

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

Observing that

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

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

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

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

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

for all η.

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

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

and

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

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

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

for all η.

F Quits and Firings

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

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

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

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

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

It is clear that the same logic would apply for $T > 2$. Suppose that the agent's contract terminates at (a potentially return-dependent) time τ, with the same "tree" structure: at each time t, there is a monotone function $\Phi_t(r_1, ..., r_t)$ such that the principal fires the agent if and only if $\Phi_t(r_1, ..., r_t) > 0$. Then, the compensation scheme has the following shape: if the agent works until τ, he receives:

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

for some constants $K_1, ..., K_T$.

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

G Proofs of Mathematical Lemmas

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Proof of Lemma 5

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Proof of Lemma 6

We use

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

When h is differentiable at x,

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

Proof of Lemma 7

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

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

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

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

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

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

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

Hence, we have:

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

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

References

[1] Di Nunno, G., B. Oksendal and F. Proske (2008): *Malliavin Calculus for Lévy Processes with Applications to Finance*, Springer Verlag

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

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