A Theory of Fair CEO Pay*

Pierre Chaigneau
Queen's University

Alex Edmans
LBS, CEPR, and ECGI

Daniel Gottlieb
LSE

November 26, 2023

Abstract

This paper studies optimal executive pay when the CEO has fairness concerns: if his wage falls below a perceived fair share of output, he suffers disutility that is increasing in the discrepancy. Fairness concerns do not lead to fair wages always being paid; instead, the firm threatens the CEO with unfair wages for low output to induce effort. The optimal contract sometimes involves performance shares: the CEO is paid a constant share of output if it is sufficiently high, but the wage drops discontinuously to zero if output falls below a threshold. Even if the incentive constraint is slack, the optimal contract features pay-for-performance, to address the CEO's fairness concerns and ensure his participation. This rationalizes pay-for-performance even if the CEO does not need effort incentives.

Keywords: Executive compensation, fairness, moral hazard.
JEL Classification: D86, G32, G34, J33.

*pierre.chaigneau@queensu.ca, aedmans@london.edu, d.gottlieb@lse.ac.uk.

For helpful comments, we thank David Dicks, Simon Gervais, Naveen Gondhi, Moqi Groen-Xu, Dirk Jenter, Wei Jiang, Sebastian Pfeil, Katya Potemkina, Martin Szydlowski, and conference participants at the Erasmus Corporate Governance Conference, Finance Theory Group, Financial Intermediation Research Society, Financial Management Association, INSEAD Finance Symposium, Northern Finance Association, and Research in Behavioral Finance Conference.

Standard executive compensation models assume that CEOs care about pay only for the consumption it enables. As a result, the marginal consumption utility of the bonus from improving performance must weakly exceed the marginal cost of effort to do so. Such models have contributed substantially to our understanding of executive compensation and inspired a stream of empirical research.

However, it is not clear that consumption utility is the only, or even the most important, driver of pay, given that CEOs are typically wealthy and most of their consumption needs are already met. Edmans, Gosling, and Jenter (2023) survey directors and investors on how they set executive contracts. Both sets of respondent highlight how pay is also driven by the need to ensure the CEO feels fairly treated, consistent with experimental evidence that agents have fairness concerns (see Fehr, Goette, and Zehnder (2009) for a survey). They also suggest that firm value is an important determinant of what the CEO views as a fair level of pay. If firm value has increased due to CEO effort, he expects to be rewarded. If firm value has increased (decreased) due to luck outside his control, he should share in this good (bad) luck. These findings echo the widely-replicated ultimatum game (e.g. Roth et al., 1991). If one party has been gifted an endowment, the other believes he should be offered a sizable share.

This paper studies optimal contracts when the CEO is motivated by both consumption utility and fairness concerns. We model the latter by specifying a perceived fair wage that is linear in the firm's output, i.e. the CEO believes that he deserves a certain percentage of output. If the CEO's wage falls below this fair wage, he suffers disutility that is linear in the discrepancy. If the wage is above the fair wage, his utility is either linear or concave in the wage. The principal is risk-neutral and both parties are protected by limited liability.

When the agent's utility is linear above the fair wage, the optimal contract involves a threshold below which the CEO is paid zero, and above which he receives the fair wage. This contradicts the intuition that fairness concerns should lead to the CEO receiving a fair wage for all output levels; instead, they mean that *un*fairness can be a powerful motivator. If output is sufficiently low that it is unlikely that the CEO has worked, he receives the most unfair possible wage of zero. Only if output exceeds a lower threshold is he paid the fair wage.

Innes (1990) showed that, with a risk-neutral agent, the optimal contract is "live-or-die" – zero if output is below a threshold, and the entire output above it. Such a contract is inefficient under fairness concerns. Even if the CEO works, output may fall below this threshold due to bad luck. If the CEO is paid zero, he perceives significant unfairness, which erodes his incentives to work. Thus, it is efficient to pay him a fair wage once output crosses a threshold. Such a contract – zero below a threshold and linear above it – resembles performance shares, which are common in reality. Standard models, such as Holmström (1979), do not predict discontinuous contracts. Innes (1990) predicts a sharp discontinuity where pay jumps from zero to the entire output, but such sharp discontinuities do not exist in reality. We predict a milder and thus more realistic discontinuity – when performance crosses a threshold, the wage jumps from zero, but not to the entire output. Intuitively, performance shares provide fair wages if performance is good and unfair wages if performance is bad, thus motivating good performance. If fairness concerns are sufficently low, there is an additional upper threshold above which the CEO receives the entire output.

When the agent's utility is concave above the fair wage, the basic features of the linear model remain robust – the payment is zero below a lower threshold, the fair wage above this threshold, and the entire output above a higher threshold. However, there is an additional fourth region, in between the regions in which the CEO receives the fair wage and the entire output. In this region, his payment exceeds the fair wage, and is generally convex in output. Intuitively, if performance is strong, the principal wishes to pay the CEO more than the fair wage. However, since the CEO is risk-averse, it is inefficient to give him the entire output.

In both the linear and nonlinear models, pay is increasing in output even when the incentive constraint is slack. Intuitively, satisfiying the participation constraint at least cost involves paying the fair wage over a range of outputs, to avoid the disutility from unfair wages. Since the fair wage is increasing in output, this generates a positive sensitivity of pay to performance. That pay is increasing in output even without an incentive constraint means that the firm can induce effort "for free". In a standard moral hazard model, implementing higher effort is always costly to the firm. In our model, since pay is increasing in output to secure participation, it automatically induces effort. Critics of high incentives argue that they should not be necessary – the CEO should be intrinsically motivated, and/or the board should monitor the CEO. Our model demonstrates that incentives may be used not to induce effort, but to retain a CEO with fairness concerns.

This paper is related to the theoretical literature on executive compensation, surveyed by Edmans and Gabaix (2016) and Edmans, Gabaix, and Jenter (2017). The vast majority of these theories feature moral hazard and only consumption utility. More closely related are CEO pay models that feature reference points. For example, De Meza and Webb (2007) and Dittmann, Maug, and Spalt (2010) study optimal CEO compensation in the presence of loss aversion. In our model, the fair wage is a reference point; the CEO is also loss-averse as his utility is steeper below the fair wage than above it. Our key innovation is that the fair wage depends on output, which leads to a very different optimal contract.[1] Some other models feature the CEO's utility depending on variables other than pay, although not output. For example, DeMarzo and Kaniel (2023) and Liu and Sun (2023) incorporate relative wealth concerns.[2]

An important literature, surveyed by Sobel (2005), studies the effect of fairness concerns on non-CEO contracts. Fehr and Schmidt (1999) explore inequity aversion, where an agent dislikes another agent receiving less than him, and dislikes even more another agent receiving more than him. In these models (as well as most experiments in this literature), subjects are ex ante symmetric, and so it makes sense for them to compare their consumption. They do not apply to a CEO setting, where the firm's objective function is shareholder value, which is orders of magnitude in excess

[1]In Akerlof and Yellen (1990), the agent's effort is reduced if his wage is below the fair wage, which is not contingent on output. In Hart and Moore (2008), the agent's reference point is outcomes permitted by the contract, rather than output. Neither model features loss aversion.

[2]We thus do not include peer firm pay as a potential reference point, as it will lead to similar results as a model of relative wealth concerns.

of CEO pay. An inequity aversion explanation for rewarding CEO for performance is that the board feels sorry for the CEO as his pay is so low, which seems at odds with real-life perceptions. (If the board represents individual shareholders rather than the firm, and the CEO always earns more than individual shareholders, inequity aversion would have no bite as shareholders would always want to lower pay as in a standard model). In our model, it is the CEO who has fairness preferences, not the board or shareholders. Moreover, the CEO is only concerned for his own utility, unlike in social preference models where agents are concerned with others' utility.

1 The Model

We consider a standard principal-agent model with one added feature: the agent (manager, "he") has fairness concerns, specified below.

At time $t = -1$, the principal (firm, "she") offers a contract to the agent. At $t = 0$, the agent privately chooses an effort level $e \in \mathbb{R}_+$ at cost $C(e)$, where $C(\cdot)$ is continuously differentiable with $C'(e) > 0$ and $C''(e) > 0$ for $e > 0$, $C'(0) = 0$, and $\lim_{e \nearrow \infty} C'(e) = \infty$. At $t = 1$, output $q \in [0, \overline{q}]$ is realized, where $\overline{q}$ may be finite or infinite, and the agent is paid a wage $w(q)$. Output is distributed according to a twice continuously differentiable density function $\phi(q|e)$ that satisfies the monotone likelihood ratio property ("MLRP"). Both the principal and agent are protected by limited liability, so that $0 \leq w(q) \leq q \ \forall q$.

Due to fairness concerns, the agent's utility $u(w, q)$ depends on both his wage w and output q, and is increasing and continuously differentiable in the former. His reservation utility is $\overline{U}$.

For each target level e^T, the principal's problem is to find the cheapest contract that induces an effort of at least e^T:

$$\min_{w(\cdot), e^*} \int_0^{\overline{q}} w(q)\phi(q|e^*)dq \tag{1}$$

$$\text{s.t.} \quad e^* \in \arg\max_e \left\{ \int_0^{\overline{q}} u(w(q), q)\phi(q|e)dq - C(e) \right\} \geq e^T \tag{2}$$

$$\int_0^{\overline{q}} u(w(q), q)\phi(q|e^*)dq - C(e^*) \geq \overline{U} \tag{3}$$

$$0 \leq w(q) \leq q \ \forall q \tag{4}$$

$$w(q) \geq w(q') \ \forall q > q' \tag{5}$$

where (2) is the incentive compatibility constraint ("IC"), (3) is the individual rationality constraint ("IR"), (4) are the limited liability constraints, and (5) is the agent's monotonicity constraint which ensures that the wage is non-decreasing in output (otherwise he would "burn" it). If the IC is slack, the effort chosen by the manager e^* will exceed e^T.

The above formulation is the standard moral hazard model in which the agent first exerts effort and then receives pay, similar to Holmström (1979) and Innes (1990), with the only departure being the agent's utility function. As a result, any deviation in the optimal contract can be attributed to fairness concerns. Another formulation would be to have a single-period model in which the agent first receives pay and then chooses effort, or a multi-period model where the agent responds to first-period pay by choosing second-period effort. Then, if offered unfair pay, he may withhold effort and reduce total surplus, as in the ultimatum game. However, such a formulation would be more ad hoc, as we would need to hard-wire the link between perceived unfairness and next-period effort, rather than fairness entering the utility function. It would also be less comparable with standard models.

Define the likelihood ratio $LR\left(q|e\right)$ as follows:

$$LR(q|e) \equiv \frac{\frac{\partial \phi}{\partial e}(q|e)}{\phi(q|e)},$$

and let q_0^e be the output for which the likelihood ratio is zero: $LR(q_0^e|e) = 0$. By MLRP and the differentiability of ϕ, q_0^e exists and is unique. To guarantee an optimal contract exists, we assume:

$$\int_0^{q_0^{e^T}} u(0,q)\frac{\partial \phi}{\partial e}(q|e^T)dq + \int_{q_0^{e^T}}^{\overline{q}} u(q,q)\frac{\partial \phi}{\partial e}(q|e^T)dq \geq C'(e^T) \tag{6}$$

This means that a contract paying the agent the minimum (zero) for outputs with negative likelihood ratios, and the maximum (the entire output) for outputs with positive likelihood ratios, will induce e^T.

Lemma 1 below derives a sufficient condition for the validity of the first-order approach ("FOA"), which allows us to replace the IC (2) by its first-order condition.[3] Let $K_e^+(q)$ and $K_e^-(q)$ denote the positive and negative parts of the second derivative of $\phi(q|e)$ with respect to effort:

$$K_e^+(q) := \max\left\{\frac{\partial^2 \phi}{\partial e^2}(q|e),\ 0\right\}, \tag{7}$$

$$K_e^-(q) := \min\left\{\frac{\partial^2 \phi}{\partial e^2}(q|e),\ 0\right\}. \tag{8}$$

Lemma 1 *(First-Order Approach): The FOA is valid if*

$$\int_0^{\overline{q}} \left(K_e^-(q)u(0,q) + K_e^+(q)u(q,q)\right)dq < C''(e) \tag{9}$$

for all $e \in \mathbb{R}_+$.

We henceforth assume that condition (9) holds.

[3] This is related to the condition for the FOA in a model with limited liability in Chaigneau, Edmans, and Gottlieb (2022). The difference is that, in Lemma 1, the utility function also depends on output.

The perceived fair wage for output q, $w^*(q)$, is given by:

$$w^*(q) \equiv \rho q, \tag{10}$$

where $\rho \in (0, 1]$ is the agent's perceived fair share. One determinant of ρ is the importance of agent effort for firm output. Edmans, Gosling, and Jenter (2023) find that "how much the CEO can affect firm performance" is the most important driver of pay variability. Another potential determinant is incentives in peer firms – the third most popular response is "the split between fixed and variable pay in peer firms."

The agent's utility function is:

$$u(w, q) = \min\left\{v(w), v(w^*(q)) + (1+\gamma)(w - w^*(q))\right\}, \tag{11}$$

where $v' > 0$, $v'' \leq 0$, and $\gamma \geq 0$. If the wage is fair ($w \geq w^*(q)$), the agent's utility is $v(w)$. If the wage is unfair ($w < w^*(q)$), the agent suffers disutility which is increasing in both the discrepancy $w - w^*(q)$ and his fairness concerns $\gamma \geq 0$. We assume $\lim_{w \to 0} v'(w) < 1 + \gamma$, so that the utility function is always steeper below the fair wage than above it. The kink at the fair wage means that the agent is loss-averse: his sensitivity to losses exceeds his sensitivity to gains. The unique feature of our fairness model is that the fair wage depends on output and is thus endogenously determined ex post, in contrast to loss aversion models where the reference point is independent of output and thus known ex ante.

A special case of (11) is where the agent is risk-neutral above the fair wage ($v(w) = w$), and so:

$$u(w, q) = w - \gamma \max\left\{w^*(q) - w, 0\right\}. \tag{12}$$

This utility function is piecewise linear with a kink at the fair wage, a slope of 1 above and a slope exceeding 1 below. This is the simplest and most transparent specification for fairness concerns, and allows us to conduct comparative statics with respect to γ and ρ.

Figure 1 displays the agent's utility as a function of w for various output realizations and two different utility functions. The Online Appendix extends the analysis to a nonlinear fair wage and a convex utility function below the fair wage, as in prospect theory; all results are qualitatively unchanged.

To simplify the analysis, we assume:

$$\int_0^{q_0^{e^T}} u(0, q) \frac{\partial \phi}{\partial e}(q|e^T) dq + \int_{q_0^{e^T}}^{\overline{q}} u(\rho q, q) \frac{\partial \phi}{\partial e}(q|e^T) dq \geq C'(e^T) \tag{13}$$

$$\int_0^{\overline{q}} u(0, q) \phi(q|0) dq - C(0) < \overline{U} \tag{14}$$

$$\int_0^{\overline{q}} u(\rho q, q) \phi(q|e^*) dq - C(e^*) \geq \overline{U}, \text{ where } e^* \text{ satisfies (2) with } w(q) = w^*(q)\ \forall q. \tag{15}$$

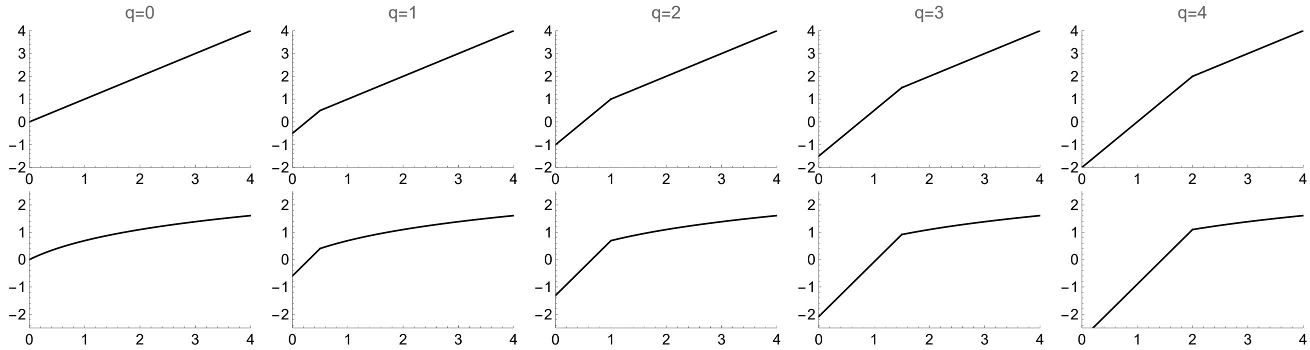

Figure 1: Function $u(w, q)$ as defined in equation (11) as a function of w for $\gamma = 1$, $\rho = \frac{1}{2}$, and output $q \in \{0, 1, 2, 3, 4\}$. Top row: $v(w) = w$. Bottom row: $v(w) = \ln(w + 1)$.

These assumptions are not crucial for our results, but reduce the number of cases we need to consider. Assumption (13) ensures that an incentive-compatible contract that elicits e^T exists even if the firm never pays more than the fair wage. Assumption (14) implies that, even if the cost of effort were zero, a contract that always pays zero would violate the IR. Assumption (15) implies that a contract that always pays the fair wage satisfies the IR.

2 Analysis

Proposition 1 studies the optimal contract when $e^T = 0$, i.e. the contract only seeks to ensure the agent's participation. Note that he still chooses effort optimally (see (2)) given the contract.

Proposition 1 *(Zero target effort level): Fix $e^T = 0$. If $\overline{U}$ is sufficiently high, the agent exerts $e^* > 0$ and the contract is characterized by $w'(q) > 0$ for some q. If $\overline{U}$ is sufficiently low, the principal implements $e^* = 0$ and the following contract is optimal:*

$$w(q) = \begin{cases} w^*(q) & \text{for } q < q_c \\ \rho q_c & \text{for } q \geq q_c \end{cases}, \tag{16}$$

where q_c is set so that the IR in equation (3) binds.

Perhaps surprisingly, an agent with a high reservation utility (threshold defined in the Appendix) chooses a strictly positive effort level even though the principal does not request any effort. Intuitively, the principal satisfies the IR by offering the fair wage for a range of outputs, and thus a wage that is increasing in output. Note that a wage increasing in output does not automatically induce effort, because even though output increases the agent's wage, it also increases the fair wage. If $\overline{U}$ is low, the principal can satisfy the agent's IR even if she only provides fair wages for a small subset of outputs. Since the fair wage is increasing in output, it is cheaper for her to provide fair wages for lower outputs, and so she pays the fair wage for $q < q_c$ and a fixed wage (which does not rise with output and is thus unfair) for $q \geq q_c$. If $q \geq q_c$, the actual wage

ρq_c is below the fair wage of ρq and so the agent suffers disutility; to avoid $q \geq q_c$, he exerts zero effort.

In contrast, if $\overline{U}$ is high, the principal needs to pay fair wages for a wider range of outputs to ensure the agent's participation. Since the fair wage is increasing in output, it induces effort as a by-product. The agent exerts $e^* > 0$ even though the principal requested $e^T = 0$; she gets effort "for free". This result is in stark contrast to the case without fairness concerns. In the standard model of Holmström (1979), eliciting higher effort is always more costly to the principal as it requires an output-contingent wage and thus inefficient risk-sharing. As a result, any effort level in $\mathbb{R}_+$ can in principle be optimal, depending on model parameters.

This is not true with fairness concerns. Providing low effort incentives either requires paying unfair wages for high outputs (which fails to satisfy the IR) or paying above the fair wage for low outputs (which is costly). Without fairness concerns, it is costly to incentivize high effort levels; with fairness concerns, it may be costly to incentivize low effort levels as doing so requires offering unfair pay. A by-product of fair pay is that it incentivizes effort, even if such incentives are unnecessary. This result may extend beyond the C-suite; for example, equity might be given to rank-and-file employees, despite their limited incentive effect, if they believe it is fair to share in the firm's fortunes.[4]

We now move to the optimal contract when the IC binds. Define $q_m^{\min}$ as the highest value that satisfies the following equation:

$$\int_0^{q_m^{\min}} u(0,q)\frac{\partial \phi}{\partial e}(q|e^T)dq + \int_{q_m^{\min}}^{\overline{q}} u(v\rho q, q)\frac{\partial \phi}{\partial e}(q|e^T)dq = C'(e^T). \tag{17}$$

If there exists a contract that implements e^T without paying the agent above the fair wage for any output, $q_m^{\min}$ is the threshold below which the payment is zero and above which it is the fair wage.

We start with the case where the agent is risk-neutral above the fair wage ($v'' = 0$).

Proposition 2 *(Binding incentive constraint, $v'' = 0$): Fix e^T sufficiently high. The principal implements $e^* = e^T$ and offers the following contract:*

$$w(q) = \begin{cases} 0 & \text{for } q < q_m \\ w^*(q) & \text{for } q \in [q_m, q_M) \\ q & \text{for } q \geq q_M \end{cases} . \tag{18}$$

Moreover:

(a) If $\gamma < \frac{LR(\overline{q}|e^T)}{LR(q_m^{min}|e^T)} - 1$ and $\overline{U}$ is sufficiently low that the IR is slack, then $LR\left(q_m|e^T\right)(1+\gamma) = LR\left(q_M|e^T\right)$ and $q_m \geq q_0^{e^T}$.

(b) If $\overline{U}$ is sufficiently high that the IR binds, then $LR(q_m|e^T)\,(1+\gamma) < LR(q_M|e^T)$.

[4] Oyer (2004) offers a different explanation for this practice – that equity-based pay ensures that wages constantly match outside opportunities.

(c) If $\gamma > \frac{LR(\overline{q}|e^T)}{LR(q_m^{min}|e^T)} - 1$ *and*

$$\overline{U} \leq -\gamma\rho \int_0^{q_m^{\min}} q\phi(q|e^T)dq + \rho \int_{q_m^{\min}}^{\overline{q}} q\phi(q|e^T)dq,$$

then $q_m = q_m^{\min}$ *and* $q_M = \overline{q}$*;*

Without fairness concerns ($\gamma = 0$), the model is similar to Innes (1990). Due to MLRP, the principal concentrates rewards on very high outputs only. Parts (a) and (b)[5] show that, regardless of whether the IR is binding, $\gamma = 0$ leads to $q_m = q_M$: the optimal contract is "live-or-die". There is a single threshold below which the agent is paid the minimum (zero) and above which he is paid the maximum (the entire output).

With fairness concerns ($\gamma > 0$), parts (a) and (b) establish that, when γ is low or the IR binds, the optimal contract has a third region: for intermediate outputs $q \in [q_m, q_M)$, the agent is paid the fair wage, and the size of the region is increasing in γ (since q_m and q_M are determined by $(1+\gamma)LR(q_m|e^*) = LR(q_M|e^*)$). The reason for this third region is that the Innes (1990) contract is suboptimal when $\gamma > 0$ on two grounds. First, it does not satisfy the IR efficiently, which is a concern if it is binding (i.e. part (b) applies). The agent receives an unfair wage (zero) for outputs below the threshold, which causes disutility. Second, it does not satisfy the IC efficiently. The agent receives an unfair wage even for some output levels that are associated with positive likelihood ratios and indicate that he has worked, reducing his incentives to do so. Since the utility function is steeper below $w^*(q)$ rather than above it, it is efficient to increase the rewards for moderately low outputs (that nevertheless have positive likelihood ratios) from 0 to $w^*(q)$, and simultaneously to reduce the rewards for moderately high outputs from q to $w^*(q)$.

While the above explains the optimal contract by starting from a model of moral hazard and adding fairness, another way to view the intuition is to start with a pure fairness model and add moral hazard. One may think that fairness concerns would lead to the agent always being paid the fair wage $w^*(q)$, to secure his participation efficiently, but such a contract does not provide incentives efficiently. Since the agent suffers disutility from an unfair wage, it is efficient to "threaten" him with the most unfair possible wage of zero for low output: fairness concerns can justify unfair wages because avoiding unfairness is a motivator. In addition, if output is sufficiently high, the agent is paid the entire output rather than the fair wage, because efficient incentive provision involves concentrating rewards in the highest likelihood ratio states.

For $q \in [q_m, q_M)$, pay-performance sensitivity ("PPS") ρ is determined by what the CEO believes to be a fair reward for performance; as explained earlier, ρ might depend on how much his effort affects output, or PPS in peer firms. In standard models with risk neutrality (e.g. Innes (1990)), PPS is 1, which is not the case for any CEO (except for 100% owner-managers). In standard models with risk aversion (e.g. Holmström (1979)), PPS is determined by a trade-off between incentives and risk aversion. However, Edmans, Gosling, and Jenter (2023) finds that

[5]Part (c) is inapplicable with $\gamma = 0$ due to MLRP and $q_m^{\min} \leq \overline{q}$.

CEO risk aversion is the least important out of seven determinants of PPS, and Becker (2006) documents a weak relationship between risk aversion and PPS.

Part (c) shows that, when γ is sufficiently high and $\overline{U}$ is sufficiently low, q_M increases all the way to $\overline{q}$. The highest region disappears, so the agent is never paid the entire output. The optimal contract thus only has two regions – zero for low outputs and the fair wage for high outputs. Intuitively, the disutility γ from unfairness is sufficiently high that the gains from paying fair wages outweigh the standard desire to pay only for very high outputs (Innes (1990)). Thus, q_M increases to the highest possible level of $\overline{q}$. Since the agent is never paid above the fair wage, incentive compatibility is achieved by setting $q_m = q_m^{\min}$ as in equation (17). The condition on $\overline{U}$ means that a contract as in equation (18) with $q_m = q_m^{\min}$ and $q_M = \overline{q}$ satisfies the IR.

This contract represents performance shares, where the agent is given shares worth ρq that are forfeited if $q < q_m$. In standard models where the likelihood ratio is a continuous function of output, such as Holmström (1979), the optimal contract is also a continuous function of output and so does not involve discontinuities. In our model, the likelihood ratio is also a continuous function of output, yet discontinuities are optimal because the threat of the most unfair possible wage incentivizes effort. In the Innes (1990) model without a monotonicity constraint, the optimal contract is discontinuous but the agent receives either nothing or everything; we are unaware of such a contract being offered in reality. To obtain more realistic contracts, Innes (1990) assumes that the principal's payoff cannot be decreasing in output, otherwise she would "burn" it or the agent would secretly inject his own funds into the company to inflate it. Innes' theory can be interpreted as either a financing model where an entrepreneur (agent) raises funds from an outside investor (principal), or a compensation model where a company (principal) offers a contract to a CEO (agent). While the two justifications for the monotonicity constraint are realistic for the financing application, they may be less relevant for the compensation application. Dispersed shareholders cannot coordinate to burn output, and while the board acts on shareholders' behalf, burning output violates directors' fiduciary duty to the company. Similarly, it would likely be illegal for the CEO to inject his own funds into the company to manipulate the stock price. Our paper obtains realistic contracts without the need for a monotonicity assumption on the principal.[6] In addition, in Innes (1990), the monotonicity constraint leads to a continuous contract; our contract features a mild discontinuity as is common in reality.

Proposition 2 shows how whether the IR binds affects the optimal contract. When the IR does not bind, as in parts (a) and (c), we have $q_m > 0$, so that the contract has a discontinuity between zero and positive payments. Paying zero for low outputs incentivizes effort, but may fail to ensure the agent's participation. Thus, when the IR binds, the contract may not have a discontinuity part (b) allows for $q_m = 0$). Overall, the IR binding is a necessary condition for pay to be a continuous function of output. The increasing use of performance shares, which do contain discontinuities,

[6] Our model does feature a monotonicity constraint on the agent (expression (5)). Standard models do not require this as monotonic contracts are automatic given the MLRP, but this need not be the case with fairness concerns. A justification is that, if the agent's payoff were decreasing in output, he would "burn" output. Innes (1990) makes a similar justification for the principal, who has less control over output than the agent.

is consistent with the participation constraint no longer binding for many CEOs – that they are willing to accept unfair pay for low output levels suggests that they are above their outside option.

Corollary 1 shows how the contract depends on the intensity of fairness concerns.

Corollary 1 *When $v'' = 0$ and the incentive constraint binds, $q_m \leq q_0^{e^T}$ and $q_M < \overline{q}$, the threshold q_m above which the manager is paid a fair wage $w^*(q)$ is decreasing in γ.*

When the IC binds and $q_m \leq q_0^{e^T}$ (the agent is only paid zero for outputs which are bad news about effort), the threshold q_m above which the manager is paid a fair wage is decreasing in γ. Intuitively, the stronger fairness concerns are, the stronger the disutility the agent suffers from receiving zero. This reinforces effort incentives, but also reduces the agent's expected utility from the contract. The principal will therefore change the thresholds q_m and q_M to reduce effort incentives and to increase the agent's expected utility. This is achieved by decreasing q_m, because this increases the range of outputs $(q_m, q_0^{e^T})$ over which the agent receives the fair wage even though they are bad news about effort. The effect on q_M is ambiguous.

The intuition for the condition $q_m \leq q_0^{e^T}$ is as follows. An increase in γ raises the disutility of zero payments. If $q_m \leq q_0^{e^T}$, then zero payments are received only if $q < q_0^{e^T}$ (i.e. for bad news outputs) and so the agent's effort incentives unambiguously rise. If $q_m > q_0^{e^T}$, the effect of increasing γ on effort incentives is ambiguous: it raises the disutility of receiving zero payments, which arise not only for all bad news outputs ($q < q_0^{e^T}$) but also for some good news outputs $q \in (q_0^{e^T}, q_m)$.

Example 1 illustrates how the optimal contract is affected by underlying parameters.

Example 1 *The agent's preferences are given by $\gamma = 1$, $\rho = \frac{1}{2}$, $C(e) = \frac{c}{10} \times e^{10}$, and $v'' = 0$. Output is lognormally distributed with parameters $e^* = 1$ and $\sigma = 1$. Optimal contract for (a) $\overline{U} = 3$ and $c = 2$. (b): $\overline{U} = 2$ and $c = 4$. (c) $\overline{U} = 2.5$ and $c = 4.75$.*

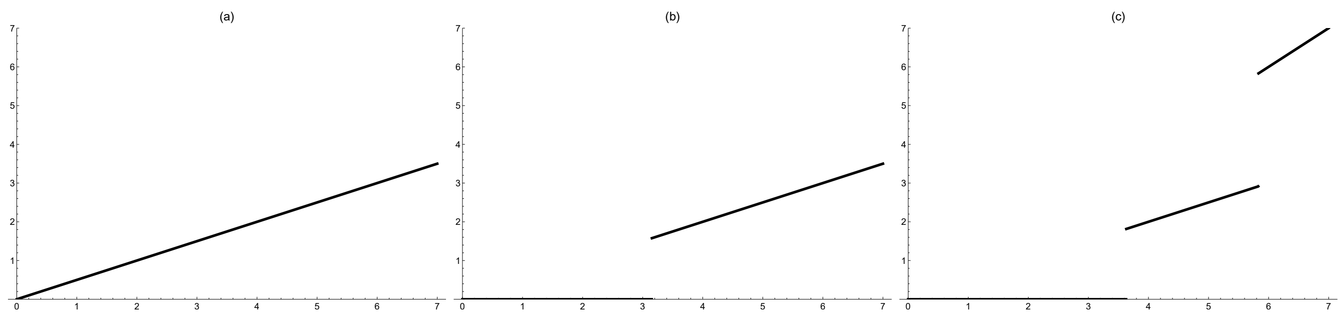

Figure 2: The contract $w(q)$ as a function of q for parameter values described in Example 1.

Proposition 3 gives the optimal contract for the case where the agent is risk-averse above the fair wage ($v'' < 0$).

Proposition 3 *(Binding incentive constraint, $v'' < 0$): Fix e^T sufficiently high. The principal implements $e^* = e^T$ and offers the following contract:*

$$w(q) = \begin{cases} 0 & \textit{for } q \in [0, q_m) \\ w^*(q) & \textit{for } q \in [q_m, q_M] \\ v'^{-1}\left(1/\left(\lambda_1 + \lambda_2 LR(q|e^*)\right)\right) & \textit{for } q \in [q_M, q_N] \\ q & \textit{for } q \in [q_N, \overline{q}] \end{cases}. \tag{19}$$

The optimal contract is now given by four regions. As with $v'' = 0$, there are three regions in which the agent is paid zero, the fair wage, and the entire output. However, there is an additional region, given by $q \in [q_M, q_N]$, where output is sufficiently high that the principal pays more than the fair wage. It is inefficient to give the entire output, since the agent exhibits diminishing marginal utility and so does not value this additional reward highly. Thus, unlike in the model with $v'' = 0$, the optimal contract is continuous at q_M. As output rises above q_M, the likelihood ratio increases further and so the actual wage exceeds the fair wage by more. The contract will generally be convex between q_M and q_N.[7] For $q > q_N$, the likelihood ratio is so high that the principal pays the entire output.

3 Conclusion

This paper studied optimal contracting under fairness preferences, where the agent's perceived fair wage depends on output. We showed that fairness concerns do not lead to the agent being paid fair wages for all output levels; in contrast, unfair wages can be effective to induce effort. When the agent's utility function is linear above the fair wage, the optimal contract involves two thresholds for output. The agent receives zero below the lower threshold, the entire output above the upper threshold, and the fair wage in between. When fairness concerns are sufficiently strong, the upper region in which the agent receives the entire output disappears, and the contract becomes performance shares. The model thus rationalizes the common usage of performance shares in reality; most other contracting theories predict continuous contracts, or extreme discontinuities where the agent's pay switches from zero to the entire output.

When the agent's utility is concave above the fair wage, the contract now involves an additional fourth region, in-between the regions in which the CEO receives the fair wage and the entire output. In this region, his payment exceeds the fair wage, and is generally convex in output.

In both models, we show that, even if the incentive constraint is slack, pay is increasing in output – by paying the agent the fair wage over a greater range of outputs, this reduces perceived unfairness and allows the participation constraint to be satisfied at least cost. As a result, the

[7]However, the contract will be concave if the likelihood ratio is concave, so that very high output is only slightly more indicative of effort, and if risk aversion is sufficiently important compared to prudence (see Chaigneau, Sahuguet and Sinclair-Desgagné, 2017). The latter condition will typically not be satisfied for CEOs who have low relative risk aversion due to their wealth.

firm can induce CEO effort "for free", potentially rationalizing why incentives are given even to intrinsically motivated agents.

This paper is a first step in modeling CEO pay under fairness preferences, using the standard model to make transparent how fairness concerns affect the optimal contract. For future research, it may be fruitful to explore the other potential determinants of the fair wage suggested by the survey of Edmans, Gosling, and Jenter (2023), such as peer firm pay in a model of multiple firms, employee pay in a model of multiple agents, or last year's pay in a dynamic model.[8]

[8]The results of Edmans, Gosling, and Jenter (2023) also suggest that shareholders, not just the CEO, also have fairness concerns. However, this may be a less promising research direction as the principal makes no decisions beyond offering the contract, and so fairness concerns do not affect effort incentives. In addition, fairness concerns for the principal are similar to a restriction on the space of contracts. Such constraints have been explored in prior work, e.g. Innes (1990).

References

[1] Akerlof, George A., and Janet L. Yellen. (1990): "The Fair Wage-Effort Hypothesis and Unemployment." *Quarterly Journal of Economics* 105, 255–283.

[2] Becker, Bo (2006): "Wealth and Executive Compensation." *Journal of Finance* 61, 379–397.

[3] Chaigneau, Pierre, Alex Edmans, and Daniel Gottlieb (2022): "How Should Performance Signals Affect Contracts?" *Review of Financial Studies* 35, 168–206.

[4] Chaigneau, Pierre, Nicolas Sahuguet, and Bernard Sinclair-Desgagné (2017): "Prudence and the Convexity of Compensation Contracts." *Economics Letters* 157, 14–16.

[5] de Meza, David and David C. Webb (2007): "Incentive Design Under Loss Aversion." *Journal of the European Economic Association* 5, 66–92.

[6] DeMarzo, Peter M. and Ron Kaniel (2023): "Contracting in Peer Networks." *Journal of Finance*, forthcoming.

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

[8] Dittmann, Ingolf, Ernst Maug, and Oliver Spalt (2010): "Sticks or Carrots? Optimal CEO Compensation when Managers Are Loss Averse." *Journal of Finance* 65, 2015–2050.

[9] Edmans, Alex and Xavier Gabaix (2016): "Executive Compensation: A Modern Primer." *Journal of Economic Literature* 54, 1232–1287.

[10] Edmans, Alex, Xavier Gabaix and Dirk Jenter (2017): "Executive Compensation: A Survey of Theory and Evidence." *Handbook of the Economics of Corporate Governance* (Benjamin E. Hermalin and Michael S. Weisbach, eds.) Elsevier/North-Holland: New York.

[11] Edmans, Alex, Tom Gosling, and Dirk Jenter (2023): "CEO Compensation: Evidence From the Field." *Journal of Financial Economics*, forthcoming.

[12] Fehr, Ernst and Klaus M. Schmidt (1999): "A Theory of Fairness, Competition, and Cooperation." *Quarterly Journal of Economics* 114, 817–868.

[13] Fehr, Ernst, Lorenz Goette, and Christian Zehnder (2009): "A Behavioral Account of the Labor Market: The Role of Fairness Concerns." *Annual Review of Economics* 1, 355–384.

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

[15] Hart, Oliver and John Moore (2008): "Contracts as Reference Points." *Quarterly Journal of Economics* 123, 1–48.

[16] Holmström, Bengt (1979): "Moral Hazard and Observability." *Bell Journal of Economics* 10, 74–91.

[17] Innes, Robert D. (1990): "Limited Liability and Incentive Contracting with Ex-Ante Action Choices." *Journal of Economic Theory* 52, 45–67.

[18] Jewitt, Ian, Ohad Kadan, and Jeroen M. Swinkels (2008): "Moral Hazard with Bounded Payments." *Journal of Economic Theory* 143, 59–82.

[19] Liu, Qi and Bo Sun (2023): "Relative Wealth Concerns, Executive Compensation, and Managerial Risk-Taking." *American Economic Journal: Microeconomics* 15, 568–598.

[20] Oyer, Paul (2004): "Why Do Firms Use Incentives That Have No Incentive Effects?" *Journal of Finance* 59, 1619–1650.

[21] Roth, Alvin E., Vesna Prasnikar, Masahiro Okuno-Fujiwara and Shmuel Zamir (1991): "Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study." *American Economic Review* 81, 1068–1095.

[22] Sobel, Joel (2005): "Interdependent Preferences and Reciprocity." *Journal of Economic Literature* 63, 392–436.

A Proofs

Proof of Lemma 1:

For a given contract $w(q)$, the effort choice problem of the agent can be written as

$$\max_e \int_0^{\overline{q}} u\left(w(q), q\right) \phi(q|e) dq - C(e).$$

The second derivative of the agent's objective function with respect to e is negative for any e if and only if:

$$\int_0^{\overline{q}} u\left(w(q), q\right) \frac{\partial^2 \phi}{\partial e^2}(q|e) dq < C''(e) \quad \forall e \in (0, \bar{e}). \tag{20}$$

With principal limited liability (see equation (4)), since the utility function increasing in w, the maximum value of u for a given q is $u\left(q, q\right)$. In addition, with agent limited liability (see equation (4)), the minimum payment is $w(q) = 0$; with a utility function increasing in w, this implies that the minimum value of u for a given q is $u\left(0, q\right)$. Therefore, for any given q:

$$u\left(w(q), q\right) \in \left[u\left(0, q\right), u\left(q, q\right)\right].$$

Using notations $K_e^+(q)$ and $K_e^-(q)$ defined in equations (7) and (8), the expression on the left-hand side ("LHS") of equation (20) can then be rewritten as:

$$\int_0^{\overline{q}} u\left(w(q), q\right) \min\left\{\frac{\partial^2 \phi}{\partial e^2}(q|e), 0\right\} dq + \int_0^{\overline{q}} u\left(w(q), q\right) \max\left\{\frac{\partial^2 \phi}{\partial e^2}(q|e), 0\right\} dq. \tag{21}$$

As established above, we have $u\left(w(q), q\right) \geq u(0, q)$ for any q, and $u\left(w(q), q\right) \leq u(q, q)$ for any q. Therefore, for any q such that $\frac{\partial^2 \phi}{\partial e^2}(q|e) \leq 0$ we have $u\left(w(q), q\right) \frac{\partial^2 \phi}{\partial e^2}(q|e) \leq u(0, q) \frac{\partial^2 \phi}{\partial e^2}(q|e)$; and for any q such that $\frac{\partial^2 \phi}{\partial e^2}(q|e) \geq 0$ we have $u\left(w(q), q\right) \frac{\partial^2 \phi}{\partial e^2}(q|e) \leq u(q, q) \frac{\partial^2 \phi}{\partial e^2}(q|e)$. Integrating over q, this implies that expression (21) is less than:

$$\int_0^{\overline{q}} \left(K_e^-(q) u(0, q) + K_e^+(q) u(q, q)\right) dq,$$

which completes the proof.

Proof of Proposition 1:

We describe the optimal contract when $e^T = 0$, i.e. the IC does not bind for any contract. In the optimization problem with a nonbinding IC, the IR for $e^* \geq 0$ must be binding. Suppose that it is not. Then, the contract that solves the optimization problem in equations (1), (4), and (5) is simply $w(q) = 0$ for any q, which gives utility $u(0, q)$ at a given q, so that the IR is not satisfied because of equation (14), a contradiction.

The relaxed optimization problem with a nonbinding IC, a binding IR, and the FOA, is:

$$\min_{w(q),e^*} \int_0^{\overline{q}} w(q)\phi(q|e^*)dq \tag{22}$$

$$\text{s.t.} \quad \int_0^{\overline{q}} u(w(q),q)\phi(q|e^*)dq - C(e^*) = \overline{U} \tag{23}$$

$$\int_0^{\overline{q}} u(w(q),q)\frac{\partial\phi}{\partial e}(q|e^*)dq = C'(e^*) \tag{24}$$

$$0 \leq w(q) \leq q \tag{25}$$

$$w(q) \geq w(q')\ \forall q > q' \tag{26}$$

At first, we hold effort constant. We then consider the effects of the contract on the effort choice, which matters because it affects the equilibrium effort e^*, and therefore the LHS of equation (23).

Lemma 2 *If the optimal contract induces $e^* = 0$, it is such that $w(q) \leq w^*(q)$ on any non-empty subinterval of* $[0, \overline{q}]$.

Proof. With $e \geq 0$ and $C'(0) = 0$, a contract $w(q)$ that induces $e^* = 0$ is such that:

$$\int_0^{\overline{q}} u(w(q),q)\frac{\partial\phi}{\partial e}(q|0)dq \leq 0. \tag{27}$$

By contradiction, suppose that the contract induces $e^* = 0$ and is such that $w(q) > w^*(q)$ on a non-empty subinterval of $[0, \overline{q}]$. Consider the following perturbation: on one of these subintervals, denoted by Q^+, decrease $w(q)$ by an arbitrarily small ϵ. On another subinterval where $w(q) < w^*(q)$, denoted by Q^-, increase $w(q)$ by ε which is such that the agent's expected utility is unchanged:

$$\begin{aligned}
&\int_{Q^+} v\left(w(q)-\epsilon\right)\phi(q|e^*)dq + \int_{Q^-} \left(v(w^*(q)) + (1+\gamma)(w(q)+\varepsilon - w^*(q))\right)\phi(q|e^*)dq \\
&= \int_{Q^+} v(w(q))\phi(q|e^*)dq + \int_{Q^-} \left(v(w^*(q)) + (1+\gamma)(w(q) - w^*(q))\right)\phi(q|e^*)dq \\
\Leftrightarrow \quad &\int_{Q^+} \left(v(w(q) - v\left(w(q)-\epsilon\right))\right)\phi(q|e^*)dq = (1+\gamma)\varepsilon\int_{Q^-}\phi(q|e^*)dq
\end{aligned}$$

where
$$\int_{Q^+} \left(v(w(q) - v\left(w(q)-\epsilon\right))\right)\phi(q|e^*)dq \leq v'(0)\epsilon\int_{Q^+}\phi(q|e^*)dq < (1+\gamma)\epsilon\int_{Q^+}\phi(q|e^*)dq \tag{28}$$

because of the assumption $v'(0) < 1+\gamma$. The implied change in expected pay is:

$$\int_{Q^+}\left(-\epsilon\right)\phi(q|e^*)dq + \int_{Q^-}\varepsilon\phi(q|e^*)dq,$$

which is strictly negative because of equation (28). ■

When the IC is nonbinding, holding effort constant (at a given $e = e^*$), the optimization program has an infinity of solutions: any monotonic contract such that $w(q) \in [0, w^*(q)]$ $\forall q$ and equation (23) holds would solve the optimization problem. Indeed, this type of contract is such that:

$$\begin{aligned} \mathbb{E}[u(w(q), q)|e^*] &= \mathbb{E}\left[v(w^*(q)) + (1+\gamma)\left(w(q) - w^*(q)\right)|e^*\right] \\ &= (1+\gamma)\mathbb{E}\left[w(q)|e^*\right] + \mathbb{E}\left[v(w^*(q))|e^*\right] - (1+\gamma)\mathbb{E}\left[w^*(q)|e^*\right], \end{aligned} \tag{29}$$

i.e., holding effort constant, the agent's expected utility only depends on $\mathbb{E}\left[w(q)|e^*\right]$. Let the threshold q_c^* be such that:

$$\int_0^{q_c^*} v(w^*(q))\phi(q|0)dq + \int_{q_c^*}^{\overline{q}} u(\rho q_c^*, q)\phi(q|0)dq = \overline{U} + C(0) \tag{30}$$

Note that this threshold exists and is unique because of the assumptions in equations (14) and (15) and because the LHS of equation (30) is continuous and strictly increasing in q_c^*.

Lemma 3 *The contract:*

$$w(q) = \begin{cases} w^*(q) & \text{if } q < q_c^* \\ \rho q_c^* & \text{if } q \geq q_c^* \end{cases}, \tag{31}$$

where q_c^* is implicitly defined in equation (30), induces $e^* = 0$ if and only if $\overline{U} \leq \overline{U}_c$. Moreover, for $\overline{U} \leq \overline{U}_c$, this contract is optimal.

The threshold $\overline{U}_c$ is the value of reservation utility $\overline{U}$ such that $e^* = 0$ solves the FOC to the agent's effort choice problem with the contract in equation (31) with a binding IR. Formally, let $\hat{q}_c$ be implicitly defined by (uniqueness is established in the proof of Lemma 3):

$$\int_0^{\hat{q}_c} v(w^*(q))\frac{\partial \phi}{\partial e}(q|0)dq + \int_{\hat{q}_c}^{\overline{q}} \left(v(w^*(q)) + (1+\gamma)\left(\rho\hat{q}_c - w^*(q)\right)\right)\frac{\partial \phi}{\partial e}(q|0)dq = C'(0).$$

Then, given $\hat{q}_c$ as in the equation above, $\overline{U}_c$ is implicitly defined by:

$$\int_0^{\hat{q}_c} v(w^*(q))\phi(q|0)dq + \int_{\hat{q}_c}^{\overline{q}} u(\rho\hat{q}_c, q)\phi(q|0)dq = \overline{U}_c + C(0).$$

Proof. Because of equation (29), with a contract such that $w(q) \in [0, w^*(q)]$ $\forall q$ and a given effort e^*, the binding IR can be rewritten as:

$$(1+\gamma)\mathbb{E}\left[w(q)|e^*\right] + \mathbb{E}\left[v(w^*(q))|e^*\right] - (1+\gamma)\mathbb{E}\left[w^*(q)|e^*\right] - C(e^*) = \overline{U}, \tag{32}$$

so that an optimal contract is such that:

$$\mathbb{E}\left[w(q)|e^*\right] = \mathbb{E}\left[w^*(q)|e^*\right] + \frac{1}{1+\gamma}\left(\overline{U} + C(e^*) - \mathbb{E}\left[v(w^*(q))|e^*\right]\right). \tag{33}$$

With a contract of the type

$$w(q) = \begin{cases} w^*(q) & \text{if } q < q_c \\ \rho q_c & \text{if } q \geq q_c \end{cases}, \tag{34}$$

there is only one q_c such that the IR in equation (23) is satisfied. Indeed:

$$\frac{\partial}{\partial q_c}\left\{\int_0^{q_c} v(w^*(q))\phi(q|e^*)dq + \int_{q_c}^{\overline{q}} \left(v(w^*(q)) + (1+\gamma)\left(\rho q_c - w^*(q)\right)\right)\phi(q|e^*)dq\right\}$$
$$= (1+\gamma)\rho\int_{q_c}^{\overline{q}} \phi(q|e^*)dq > 0. \tag{35}$$

With a contract as in equation (34), an increase in q_c leads to an increase in the LHS of the IC in equation (24):

$$\frac{\partial}{\partial q_c}\left\{\int_0^{q_c} v(w^*(q))\frac{\partial\phi}{\partial e}(q|e^*)dq + \int_{q_c}^{\overline{q}} \left(v(w^*(q)) + (1+\gamma)\left(\rho q_c - w^*(q)\right)\right)\frac{\partial\phi}{\partial e}(q|e^*)dq\right\}$$
$$= (1+\gamma)\rho\int_{q_c}^{\overline{q}} \frac{\partial\phi}{\partial e}(q|e^*)dq > 0,$$

where the inequality follows from the definition of a density function and MLRP. For $q_c = 0$, the LHS of the IC is strictly negative. By a continuity argument, it follows that $e^* = 0$ when q_c is sufficiently low. Moreover, the agent's expected utility on the LHS of equation (23) is strictly increasing in q_c (see equation (35)); for equation (23) to be satisfied, the LHS of the equation must be strictly increasing in q_c. Therefore, with a contract as in equation (34), q_c is strictly increasing in $\overline{U}$. Since an increase in q_c leads to an increase in the LHS of the IC in equation (24) and equation (27) must be satisfied for $e^* = 0$ to be the effort level optimally chosen by the agent, it follows that the contract in equation (34) induces $e^* = 0$ if and only if $\overline{U}$ is sufficiently low: $\overline{U} \leq \overline{U}_c$. ■

Lemma 4 *Effort $e^* = 0$ can be induced by a feasible contract such that $w(q) \leq w^*(q)$ if and only if it can be induced by a contract as in equation (31).*

Proof. We will show that a contract as in equation (31) minimizes the LHS of IC conditional on $e^* = 0$ and the constraints on contracting. Given $w(q) \leq w^*(q)$ and equation (29), $\mathbb{E}[u(w(q))|e^*] = \overline{U}$ is equivalent to $\mathbb{E}[w(q)|e^*] = \overline{W}$ for a given e^* (here $e^* = 0$) and some $\overline{W}$, and $u(w, q)$ is linear in w. Thus, the effort minimizing contract solves:

$$\min_{w(q)} \int w(q) \frac{\partial \phi}{\partial e}(q|0) dq \qquad \text{s.t.} \qquad \begin{aligned} &\mathbb{E}[w(q)|e^* = 0] = \overline{W} \\ &0 \leq w(q) \leq w^*(q) \\ &\dot{w}(q) \geq 0 \end{aligned}$$

Let $x(q) \equiv \dot{w}(q)$. The Hamiltonian and Lagrangian are:

$$\mathcal{H} = -w(q)\frac{\partial \phi}{\partial e}(q|0) + \theta w(q)\phi(q|0) + \lambda(q)x(q) \tag{36}$$

$$\mathcal{L} = \mathcal{H} + \mu(q)x(q) + \nu(q)w(q) + \omega(w^*(q) - w(q)) \tag{37}$$

The optimality condition with respect to the control variable x is:

$$\frac{\partial \mathcal{L}}{\partial x} = 0 \quad \Leftrightarrow \quad \lambda(q) = -\mu(q) \tag{38}$$

By complementary slackness, $\mu(q) \geq 0$, and $\mu(q) = 0$ if $x(q) > 0$. Thus, $\lambda(q) \leq 0$, and $\lambda(q) = 0$ if $x(q) > 0$. The equation of motion for the costate variable is:

$$\dot{\lambda}(q) = -\frac{\partial \mathcal{L}}{\partial w} \quad \Leftrightarrow \quad \dot{\lambda}(q) = \frac{\partial \phi}{\partial e}(q|0) - \theta\phi(q|0) - \nu(q) + \omega(q) \tag{39}$$

The transversality condition is $\lambda(\overline{q}) = 0$. By complementary slackness, if $w(q) \in (0, w^*(q))$ and $\dot{w}(q) > 0$ then:

$$\dot{\lambda}(q) = \frac{\partial \phi}{\partial e}(q|0) - \theta\phi(q|0) \tag{40}$$

Moreover, as already established, $\lambda(q) = 0$ if $x(q) > 0 \;\; \Leftrightarrow \;\; \dot{w}(q) > 0$, which implies $\dot{\lambda}(q) = 0$. Thus, with equation (40), $\dot{\lambda}(q) = 0$ would imply that the likelihood ratio $\frac{\frac{\partial \phi}{\partial e}(q|0)}{\phi(q|0)}$ is constant, which is impossible by MLRP. In sum, we cannot have $w(q) \in (0, w^*(q))$ and $\dot{w}(q) > 0$ on any non-empty subinterval. This implies that, on any non-empty subinterval, either $w(q) = w^*(q)$ or $\dot{w}(q) = 0$ (the latter allows for $w(q) = 0$).

Finally, we have to show that $w(q) = w^*(q)$ for q less than a threshold and $\dot{w}(q) = 0$ for q higher than this threshold. By contradiction, suppose that the contract is such that $\dot{w}(q) = 0$ (and $w(q) \leq w^*(q)$) for $q \in Q_1$ and $w(q') = w^*(q')$ for $q' \in Q_2$ where Q_1 and Q_2 are such that $q < q'$. Note that the utility function is linear in w for $w(q) \leq w^*(q)$. Standard arguments show that it is possible to perturb the contract by increasing payments on Q_1 and decreasing payments on Q_2 to maintain the same expected payment $\int w(q)\phi(q|0)dq$ while reducing effort incentives $\int w(q)\frac{\partial \phi}{\partial e}(q|0)dq$ because of MLRP. ■

Lemma 5 *If $e^T = 0$ and $\overline{U} \leq \overline{U}_c$, then the principal implements $e^* = 0$.*

Proof. First, in this case, an optimal contract cannot be such that $w(q) > w^*(q)$ on a non-empty subinterval. As already established, a contract as in equation (31) with q_c such that IR binds induces $e^* = 0$ when $\overline{U} \leq \overline{U}_c$, i.e. in this case there exists a contract such that $w(q) \leq w^*(q)$ for any q that induces $e^* = 0$. Now consider a contract $w^A(q)$ such that $w^A(q) > w^*(q)$ on a non-empty subinterval. There are two cases. If this contract induces $e^* = 0$, then this contract is dominated (see Lemma 2). If this contract $w^A(q)$ induces $e^* > 0$, holding effort constant at $e^* = 0$ for now, we have:

$$\mathbb{E}[w^A(q)|e^* = 0] > \mathbb{E}[w(q)|e^* = 0], \tag{41}$$

where the contract $w(q)$ is as in equation (31), so that it induces $e^* = 0$. The inequality in equation (41) can be established as in the proof of Lemma 2. In addition, for any feasible contract $w^A(q)$ (is which nondecreasing in q for all q given the constraint in equation (26)), we have:

$$\mathbb{E}[w^A(q)|e^* > 0] \geq \mathbb{E}[w^A(q)|e^* = 0], \tag{42}$$

which follows because MLRP implies FOSD. In sum, the contract $w^A(q)$ is dominated.

Second, consider a contract $w(q)$ such that $w(q) \leq w^*(q)$ for all q that does not induce $e^* = 0$, and which is such that IR is satisfied as an equality as in equation (23) (we already showed that IR must be binding). As established in equation (29), holding effort constant, the agent's expected utility for this type of contract only depends on the expected cost of the contract, $\mathbb{E}[w(q)|e^*]$, which is as in equation (33). Moreover, for any feasible contract, a reduction in effort leads to a lower expected cost since MLRP implies FOSD (see equation (42)). Therefore, a contract such that $w(q) \leq w^*(q)$ that induces $e^* = 0$ dominates a contract such that $w(q) \leq w^*(q)$ that induces $e^* > 0$. Finally, a contract as in equation (31) with q_c such that IR binds induces $e^* = 0$ when $\overline{U} \leq \overline{U}_c$, and this contract is optimal for $e^* = 0$ (see Lemma 3). ■

In sum, if $\overline{U} \leq \overline{U}_c$, then $e^* = 0$ and an optimal contract is as in equation (31) (see Lemmas 3 and 4).

Conversely, if $\overline{U} > \overline{U}_c$, Lemmas 3 and 4 imply that any contract that induces $e^* = 0$ must be such that $w(q) > w^*(q)$ on non-empty subinterval(s). Moreover, some of these subintervals must be in $[0, q_0^0]$: if $\overline{U} > \overline{U}_c$ and $w(q) \leq w^*(q)$ for all $q < q_0^0$, then $e^* > 0$ because of Lemmas 3 and 4 and because the LHS of the IC conditional on $e^* = 0$ writes as:

$$\int_0^{\overline{q}} u(w(q), q) \frac{\partial \phi}{\partial e}(q|0) dq, \tag{43}$$

with $\frac{\partial \phi}{\partial e}(q|0) < 0$ if and only if $q < q_0^0$ by MLRP.

We still consider the case $\overline{U} > \overline{U}_c$. By contradiction, suppose that the optimal contract is such that $w(q) > w^*(q)$ on a non-empty set of subintervals whose union is denoted by Q^+, and $e^* = 0$. We distinguish the following subintervals: Q_a^+ which include q such that $w(q) > w^*(q)$

and $q < q_0^0$; Q_b^+ which include q such that $w(q) > w^*(q)$ and $q > q_0^0$; Q^- which include q such that $w(q) \leq w^*(q)$; let $Q^+ = Q_a^+ \cup Q_b^+$. Note that any optimal contract such that $e^* = 0$ must be such that Q_b^+ is empty, so that we henceforth refer to Q_a^+ simply as Q^+. Consider a perturbation similar to the proof of Lemma (2). Specifically, decrease $w(q)$ by an arbitrarily small ϵ on Q^+, and increase $w(q)$ by an arbitrarily small ε on Q^-, where ε is such that the agent remains at his reservation level of utility. There are two cases. First, if the perturbation does not change the optimal level of effort, which remains $e^* = 0$, then standard arguments (see below with $\frac{de^*}{d\epsilon} = 0$) show that the expected cost of the contract decreases. Second, consider the case when the perturbation marginally increases the optimal level of effort from $e^* = 0$. The change in expected utility from a marginal change in the optimal level of effort for a given ϵ perturbation is:

$$\epsilon \frac{de^*}{d\epsilon}\left(\int_{Q^+} v(w(q))\frac{\partial \phi}{\partial e}(q|0)dq + \int_{Q^-}\left(v(w^*(q)) + (1+\gamma)(w(q) - w^*(q))\right)\frac{\partial \phi}{\partial e}(q|0)dq - C'(0)\right)$$

By the envelope theorem, since the agent chooses effort optimally, the expression in this equation is equal to zero. By construction, ϵ and ε are such that the change in the agent's expected utility is zero:

$$\begin{aligned}
& \int_{Q^+} v\left(w(q) - \epsilon\right)\phi(q|e^*)dq + \int_{Q^-}\left(v(w^*(q)) + (1+\gamma)(w(q) + \varepsilon - w^*(q))\right)\phi(q|e^*)dq \\
& = \int_{Q^+} v(w(q))\phi(q|e^*)dq + \int_{Q^-}\left(v(w^*(q)) + (1+\gamma)(w(q) - w^*(q))\right)\phi(q|e^*)dq \\
\Leftrightarrow \quad & \int_{Q^+}\left(v(w(q) - v\left(w(q) - \epsilon\right))\right)\phi(q|e^*)dq = (1+\gamma)\varepsilon\int_{Q^-}\phi(q|e^*)dq \\
\text{where} \quad & \int_{Q^+}\left(v(w(q) - v\left(w(q) - \epsilon\right))\right)\phi(q|e^*)dq \leq v'(0)\epsilon\int_{Q^+}\phi(q|e^*)dq < (1+\gamma)\epsilon\int_{Q^+}\phi(q|e^*)dq
\end{aligned} \tag{44}$$

This implies that the change in expected pay, which is:

$$\int_{Q^+} -\epsilon\phi(q|0)dq + \int_{Q^-}\varepsilon\phi(q|0)dq, \tag{45}$$

is strictly negative. This has two implications. First, if $\overline{U} > \overline{U}_c$, given $w^*(0) = 0$, $w^{*\prime}(q) > 0$, the monotonicity constraint, and the assumption in equation (14), we have $w'(q) > 0$ for some q. Second, if $\overline{U} > \overline{U}_c$, then $e^* > 0$.

Proof of Proposition 2:

When the condition from Lemma 1 holds so that the FOA applies, a binding IC can be rewritten as:

$$\int_0^{\overline{q}} u(w(q), q)\frac{\partial \phi}{\partial e}(q|e^*)dq = C'(e^*), \tag{46}$$

with $e^* = e^T$ since the IC binds.

Lemma 6 *The optimal contract is such that $w(q) \notin (0, w^*(q))$ for any q.*

Proof. This proof is by contradiction. Suppose that for some q we have $w(q) \in (0, w^*(q))$. Consider any given initial incentive-compatible contract and the following perturbation for any $q > q'$ in this subinterval, increase $w(q)$ by $\epsilon/\phi(q|e^*)$, and decrease $w(q')$ by $\epsilon/\phi(q'|e^*)$, where ϵ is positive and arbitrarily small. By construction, for a given effort this perturbation does not change the principal's or the agent's objective function (note that the agent's objective function is linear in w for any w and q such that $w > w^*(q)$). Now consider the effect on the LHS of the IC in equation (46). With $w(q) \in (0, w^*(q))$ and $w(q') \in (0, w^*(q'))$, the change in the LHS of the IC is:

$$(1+\gamma)\epsilon \left(LR(q|e^*) - LR(q'|e^*) \right),$$

which is strictly positive by MLRP. Since the LHS of the IC increases and the IC is binding, standard arguments show that it is then possible to construct a contract that leaves the LHS of the IC and IR unchanged compared to the initial contract and reduces the cost of the contract to the principal, which establishes that the initial contract was suboptimal. This rules out any contract such that $w(q) \in (0, w^*(q))$ for any q. ■

Case 1: $v'' = 0$.

In this case, we have $v(w) = a + bw$ for $b > 0$. We normalize $v(w) = w$ (as in equation (12)).

The first step in this part of the proof establishes that a contract as described in equation (18) is optimal. To this end, we rely on the agent's monotonicity constraint in equation (5) and on Lemma 7 below.

Lemma 7 *Let the utility function be as in equation (12) and suppose that the IC is binding. The optimal contract is such that $w(q) \notin (w^*(q), q)$ for any q.*

Proof. This proof is by contradiction. Suppose that for some q we have $w(q) \in (w^*(q), q)$. Consider any given initial incentive-compatible contract and the following perturbation for any $q > q'$ in this subinterval, increase $w(q)$ by $\epsilon/\phi(q|e^*)$, and decrease $w(q')$ by $\epsilon/\phi(q'|e^*)$, where ϵ is positive and arbitrarily small. By construction, for a given effort this perturbation does not change the principal's or the agent's objective function (note that the agent's objective function is linear in w for any w and q such that $w > w^*(q)$). Now consider the effect on the LHS of the IC in equation (46). With $w(q) \in (w^*(q), q)$ and $w(q') \in (w^*(q'), q')$, the change in the LHS of the IC is:

$$\epsilon \left(LR(q|e^*) - LR(q'|e^*) \right),$$

which is strictly positive by MLRP. Since the LHS of the IC increases and the IC is binding, standard arguments show that it is then possible to construct a contract that leaves the LHS of the IC and IR unchanged compared to the initial contract and reduces the cost of the contract to the principal, which establishes that the initial contract was suboptimal. This rules out any contract such that $w(q) \in (w^*(q), q)$ for any q. ■

The second step of the proof establishes the values of q_m and q_M for a given effort e^* to be induced.

The relaxed optimization problem with $q_m \in [0, \overline{q}]$ and $q_M \in [q_m, \overline{q}]$ is:

$$\min_{q_m, q_M} \int_0^{\overline{q}} w(q)\phi(q|e^*)dq \tag{47}$$

$$\text{s.t.} \quad \int_0^{\overline{q}} u(w(q), q)\frac{\partial \phi}{\partial e}(q|e^*)dq = C'(e^*) \tag{48}$$

$$\int_0^{\overline{q}} u(w(q), q)\phi(q|e^*)dq - C(e^*) \geq \overline{U} \tag{49}$$

$$w(q) = \begin{cases} 0 & \text{for } q < q_m \\ w^*(q) & \text{for } q \in [q_m, q_M] \\ q & \text{for } q > q_M \end{cases} \tag{50}$$

With the utility function defined in equation (12), this can be rewritten as, for $q_m \in [0, \overline{q}]$ and $q_M \in [q_m, \overline{q}]$:

$$\min_{q_m, q_M} \int_{q_m}^{q_M} \rho q\phi(q|e^*)dq + \int_{q_M}^{\overline{q}} q\phi(q|e^*)dq \tag{51}$$

$$\text{s.t.} \int_0^{q_m} (-\gamma\rho q)\frac{\partial}{\partial e}\phi(q|e^*)dq + \int_{q_m}^{q_M} \rho q\frac{\partial \phi}{\partial e}(q|e^*)dq + \int_{q_M}^{\overline{q}} q\frac{\partial \phi}{\partial e}\phi(q|e^*)dq = C'(e^*) \tag{52}$$

$$\int_0^{q_m} (-\gamma\rho q)\,\phi(q|e^*)dq + \int_{q_m}^{q_M} \rho q\phi(q|e^*)dq + \int_{q_M}^{\overline{q}} q\phi(q|e^*)dq - C(e^*) \geq \overline{U} \tag{53}$$

Denote by η_{IC} and η_{IR} the Lagrange multipliers associated with the constraints in equations (52) and (53), respectively. The FOC for an interior solution are:

$$\begin{aligned} -\rho q_m\phi(q_m|e^*) - \eta_{IC}\left(-\gamma\rho q_m\frac{\partial \phi}{\partial e}(q_m|e^*) - \rho q_m\frac{\partial}{\partial e}\phi(q_m|e^*)\right) \\ -\eta_{IR}\left(-\gamma\rho q_m\phi(q_m|e^*) - \rho q_m\phi(q_m|e^*)\right) = 0 \end{aligned} \tag{54}$$

$$\begin{aligned} \rho q_M\phi(q_M|e^*) - q_M\phi(q_M|e^*) - \eta_{IC}\left(\rho q_M\frac{\partial \phi}{\partial e}(q_M|e^*) - q_M\frac{\partial \phi}{\partial e}(q_M|e^*)\right) \\ -\eta_{IR}\left(\rho q_M\phi(q_M|e^*) - q_M\phi(q_M|e^*)\right) = 0 \end{aligned} \tag{55}$$

which for $q_m \neq 0$ and $q_M \neq 0$ is equivalent to:

$$-1 + \eta_{IC}\frac{\frac{\partial \phi}{\partial e}(q_m|e^*)}{\phi(q_m|e^*)}(1+\gamma) + \eta_{IR}(1+\gamma) = 0 \tag{56}$$

$$-1 + \eta_{IC}\frac{\frac{\partial \phi}{\partial e}(q_M|e^*)}{\phi(q_M|e^*)} + \eta_{IR} = 0 \tag{57}$$

The optimal value of q_m is generically not described by a corner solution. We have $q_m = 0$ in a nongeneric case: when equation (15) is satisfied as an equality at $e^* = e^T$. Now suppose that equation (15) is not satisfied as an equality at $e^* = e^T$, i.e. it is satisfied as a strict inequality. This implies that the IR is nonbinding when $q_m = 0$ (indeed, it is nonbinding for $q_m = 0$ and $q_M = \overline{q}$, and the LHS of the IR in equation (53) is decreasing in q_M). Moreover, a contract with $q_m = 0$ does not provide incentives at the minimum cost, since increasing q_m would increase the LHS of the IC in equation (52) while reducing the cost of the contract in equation (51). Finally, since the IR is nonbinding at $q_m = 0$ and its LHS is continuously differentiable in q_m, the increase in q_m can be small enough that the new contract still satisfies IR. In sum, if equation (15) is not satisfied as an equality at $e^* = e^T$, then we cannot have $q_m = 0$ at the optimal contract.

Likewise, we cannot have $q_m = \overline{q}$, which would imply $q_M = \overline{q}$, at the optimal contract. Indeed, this would violate the IC in equation (52) since the LHS would then be negative and the RHS positive; this would also violate the IR in equation (53) according to equation (14).

Thus, the optimal value of q_m is generically given by the first-order condition in equation (56), which can be rearranged as:

$$LR(q_m|e^*) = \frac{1}{\eta_{IC}}\left(\frac{1}{1+\gamma} - \eta_{IR}\right),$$

where $\eta_{IC} \geq 0$ and $\eta_{IR} \geq 0$.

There are two cases.

Nonbinding IR. In the optimization problem with a nonbinding IR, the IC for $e^T > 0$ must be binding. Suppose that it is not. Then, the contract that solves the optimization problem in equations (1), (4), and (5) is simply $w(q) = 0$ for any q, so that $u(0, q) = -\gamma \max\{\rho q, 0\} = -\gamma\rho q$ for any $q \in [0, \overline{q}]$, and:

$$\int_0^{\overline{q}} u(0,q)\frac{\partial \phi}{\partial e}(q|e)dq = -\gamma\rho\int_0^{\overline{q}} q\frac{\partial \phi}{\partial e}(q|e)dq < 0 < C'(e),$$

for any $e > 0$, i.e. the IC is not satisfied, a contradiction.

If the optimal values of q_m and q_M are interior solutions, equations (56) and (57) with $\eta_{IR} = 0$ (nonbinding IR) and $\eta_{IC} > 0$ (binding IC) immediately give:

$$\frac{\frac{\partial\phi}{\partial e}(q_m|e^*)}{\phi(q_m|e^*)}(1+\gamma) = \frac{\frac{\partial\phi}{\partial e}(q_M|e^*)}{\phi(q_M|e^*)}. \tag{58}$$

With a nonbinding IR, we establish that $q_m \geq q_0^{e^*}$, where $e^* = e^T$. Consider any given initial compensation contract such that $q_m < q_0^{e^*}$ and the following perturbation: increase q_m by an arbitrarily small amount. This perturbation increases the LHS of the IC and reduces the cost of the contract to the principal. Standard arguments show that it is then possible to construct a contract that leaves the LHS of the IC unchanged compared to the initial contract and reduces the cost of the contract to the principal, which establishes that the initial contract was suboptimal.

Denote the subset of values of $\{q_m, q_M\}$ that satisfy the IC by $\mathcal{Q}^{IC}$, and denote the values of $\{q_m, q_M\}$ in this subset by $\{q_m^{IC}, q_M^{IC}\}$. Let q_M^{IC} be a function of q_m^{IC}. This is a continuous function by the implicit function theorem since the LHS of the IC in equation (48) is continuously differentiable in q_m and q_M, and the product of continuous functions is continuous.

Totally differentiating the LHS of the IC with respect to q_m^{IC} and taking into account the effect on q_M^{IC} so that the LHS of the IC remains unchanged gives:

$$\begin{aligned}
\frac{d}{dq_m^{IC}}\int_0^{\overline{q}} u(w(q),q)\frac{\partial\phi}{\partial e}(q|e^*)dq &= \left(u(0,q_m^{IC}) - u(w^*(q_m^{IC}),q_m^{IC})\right)\frac{\partial\phi}{\partial e}(q_m^{IC}|e^*) \\
&\quad + \left(\left(u(w^*(q_M^{IC}),q_M^{IC}) - u(q_M^{IC},q_M^{IC})\right)\frac{\partial\phi}{\partial e}(q_M^{IC}|e^*)\right)\frac{dq_M^{IC}}{dq_m^{IC}} \\
&= -(1+\gamma)w^*(q_m^{IC})\frac{\partial\phi}{\partial e}(q_m^{IC}|e^*) - (q_M^{IC} - w^*(q_M^{IC}))\frac{\partial}{\partial e}\phi(q_M^{IC}|e^*)\frac{dq_M^{IC}}{dq_m^{IC}} = 0 \\
\Leftrightarrow \quad \frac{dq_M^{IC}}{dq_m^{IC}} &= -\frac{(1+\gamma)w^*(q_m^{IC})}{q_M^{IC} - w^*(q_M^{IC})}\frac{\frac{\partial}{\partial e}\phi(q_m^{IC}|e^*)}{\frac{\partial\phi}{\partial e}(q_M^{IC}|e^*)}, \qquad (59)
\end{aligned}$$

where both the numerator and the denominator of the second fraction on the RHS are positive since $q_0^{e^*} \leq q_m \leq q_M$.

Now consider the subset $\mathcal{Q}^c$ of values of $\{q_m, q_M\}$, denoted by $\{q_m^c, q_M^c\}$, that leaves the expected cost of the contract in equation (47) unchanged for the principal. By construction:

$$\begin{aligned}
\frac{d}{dq_m^c}\int_0^{\overline{q}} w(q)\phi(q|e^*)dq = -w^*(q_m^c)\phi(q_m^c|e^*) - (q_M^c - w^*(q_M^c))\phi(q_M^c|e^*)\frac{dq_M^c}{dq_m^c} &= 0 \\
\Leftrightarrow \quad \frac{dq_M^c}{dq_m^c} = -\frac{w^*(q_m^c)}{q_M^c - w^*(q_M^c)}\frac{\phi(q_m^c|e^*)}{\phi(q_M^c|e^*)}&. \qquad (60)
\end{aligned}$$

Because of MLRP, for $q_0^{e^*} \leq q_m \leq q_M$, we have:

$$\frac{\frac{\partial\phi}{\partial e}(q_m|e^*)}{\phi(q_m|e^*)} \leq \frac{\frac{\partial\phi}{\partial e}(q_M|e^*)}{\phi(q_M|e^*)} \quad \Leftrightarrow \quad \frac{\frac{\partial}{\partial e}\phi(q_m|e^*)}{\frac{\partial\phi}{\partial e}(q_M|e^*)} \leq \frac{\phi(q_m|e^*)}{\phi(q_M|e^*)}, \qquad (61)$$

with strict inequalities for $q_M > q_m$.

For any given element in $\mathcal{Q}^{IC}$, there are two possible cases:

1) For values of q_m^{IC} and q_M^{IC} such that $(1+\gamma)\frac{\frac{\partial\phi}{\partial e}(q_m^{IC}|e^*)}{\phi(q_m^{IC}|e^*)} < \frac{\frac{\partial\phi}{\partial e}(q_M^{IC}|e^*)}{\phi(q_M^{IC}|e^*)}$ and $q_m \geq q_0^{e^*}$, a marginal increase in q_m and associated decrease in q_M (since $\frac{dq_M^{IC}}{dq_m^{IC}} < 0$) that satisfies incentive compatibility as in equation (59) results in a lower cost to the principal because of equations (60) and (61).

2) For values of q_m^{IC} and q_M^{IC} such that $(1+\gamma)\frac{\frac{\partial\phi}{\partial e}(q_m^{IC}|e^*)}{\phi(q_m^{IC}|e^*)} > \frac{\frac{\partial\phi}{\partial e}(q_M^{IC}|e^*)}{\phi(q_M^{IC}|e^*)}$ and $q_m \geq q_0^{e^*}$, a marginal increase in q_m and associated decrease in q_M (since $\frac{dq_M^{IC}}{dq_m^{IC}} < 0$) that satisfies incentive compatibility as in equation (59) results in a higher cost to the principal because of equations (60)

and (61).

Consider the smallest value for q_m^{IC} and corresponding highest value for q_M^{IC} in the subset $\mathcal{Q}^{IC}$, and denote them by $q_m^{\min}$ and $q_M^{\max}$. We can show by construction that $q_M^{\max} = \overline{q}$: according to equations (13), an incentive-compatible contract such that $q_m \geq q_0^{e^*}$ and $q_M = \overline{q}$ exists; by definition of $\mathcal{Q}^{IC}$ and $q_M^{\max}$, this means that $q_M^{\max} = \overline{q}$. Since IR is nonbinding and the cost of a contract is decreasing in q_m, all else equal, $q_m^{\min}$ is implicitly defined by incentive compatibility with $q_M^{\max} = \overline{q}$ in equation (17). From equations (13) and (17), we have $q_m^{\min} \geq q_0^{e^*}$. There are two cases.

First, if $(1+\gamma)\frac{\frac{\partial\phi}{\partial e}(q_m^{\min}|e^*)}{\phi(q_m^{\min}|e^*)} > \frac{\frac{\partial\phi}{\partial e}(\overline{q}|e^*)}{\phi(\overline{q}|e^*)}$, then due to MLRP and $\frac{dq_M^{IC}}{dq_m^{IC}} < 0$, for any element of $\mathcal{Q}^{IC}$, we have $(1+\gamma)\frac{\frac{\partial\phi}{\partial e}(q_m^{IC}|e^*)}{\phi(q_m^{IC}|e^*)} > \frac{\frac{\partial\phi}{\partial e}(q_M^{IC}|e^*)}{\phi(q_M^{IC}|e^*)}$, so that case 2) described above is relevant for any element of $\mathcal{Q}^{IC}$. Therefore, the optimal values of q_m and q_M are respectively $q_m^{\min}$ and $\overline{q}$. That is:

$$w(q) = \begin{cases} 0 & \text{for } q \in [0, q_m^{\min}) \\ w^*(q) & \text{for } q \in [q_m^{\min}, \overline{q}] \end{cases}, \tag{62}$$

where $q_m^{\min}$ is defined in equation (17).

Second, if $(1+\gamma)\frac{\frac{\partial\phi}{\partial e}(q_m^{\min}|e^*)}{\phi(q_m^{\min}|e^*)} < \frac{\frac{\partial\phi}{\partial e}(\overline{q}|e^*)}{\phi(\overline{q}|e^*)}$, then for elements in the subset $\mathcal{Q}^{IC}$, for low enough values of q_m^{IC} and high enough values of q_M^{IC}, case 1) described above is relevant. Moreover, since $\gamma > 0$ and the likelihood ratio $LR(q|e)$ is continuous in q by assumption, for elements in the subset $\mathcal{Q}^{IC}$, for high enough values of q_m^{IC} and low enough values of q_M^{IC} (since $\frac{dq_M^{IC}}{dq_m^{IC}} < 0$), case 2) described above is relevant. In sum, the optimal values of q_m and q_M belong to the subset $\mathcal{Q}^{IC}$ and satisfy the following equation:

$$(1+\gamma)\frac{\frac{\partial\phi}{\partial e}(q_m|e^*)}{\phi(q_m|e^*)} = \frac{\frac{\partial\phi}{\partial e}(q_M|e^*)}{\phi(q_M|e^*)}. \tag{63}$$

Binding IR. When both the IC and IR are binding, q_m and q_M must satisfy:

$$-\gamma\rho\int_0^{q_m} q\frac{\partial\phi}{\partial e}(q|e^*)dq + \rho\int_{q_m}^{q_M} q\frac{\partial\phi}{\partial e}(q|e^*)dq + \int_{q_M}^{\overline{q}} q\frac{\partial\phi}{\partial e}(q|e^*)dq = C'(e^*) \tag{64}$$

$$-\gamma\rho\int_0^{q_m} q\phi(q|e^*)dq + \rho\int_{q_m}^{q_M} q\phi(q|e^*)dq + \int_{q_M}^{\overline{q}} q\phi(q|e^*)dq - C(e^*) = \overline{U} \tag{65}$$

We also know that the optimal value of q_m is generically an interior solution, so we have three cases.

1. First, if the optimal values of q_m and q_M are interior solutions, equations (56) and (57) with $\eta_{IR} > 0$ and $\eta_{IC} > 0$ immediately give:

$$\frac{\frac{\partial\phi}{\partial e}(q_m|e^*)}{\phi(q_m|e^*)}(1+\gamma) + \frac{\eta_{IR}}{\eta_{IC}}\gamma = \frac{\frac{\partial\phi}{\partial e}(q_M|e^*)}{\phi(q_M|e^*)} \tag{66}$$

Because of MLRP, the LHS of equation (66) is strictly increasing in q_m, and the RHS is strictly increasing in q_M. Thus, for any pair $\{q_m, q_M\}$ that satisfy this equation, q_M is strictly increasing in q_m.

2. If $q_M = q_m$, then q_m must satisfy:

$$-\gamma\rho \int_0^{q_m} q \frac{\partial \phi}{\partial e}(q|e^*)dq + \int_{q_m}^{\overline{q}} q \frac{\partial \phi}{\partial e}(q|e^*)dq = C'(e^*) \tag{67}$$

$$-\gamma\rho \int_0^{q_m} q\phi(q|e^*)dq + \int_{q_m}^{\overline{q}} q\phi(q|e^*)dq - C(e^*) = \overline{U} \tag{68}$$

The LHS of the IR in equation (68) is strictly decreasing in q_m. Thus, there exists at most one value of q_m such that equation (68) holds, and this value is strictly decreasing in $\overline{U}$. The derivative of the LHS of IC in equation (67) with respect to q_m is $q_m \frac{\partial \phi}{\partial e}(q_m|e^*)\left(-\gamma\rho - 1\right)$, which by MLRP and definition of $q_0^{e^*}$ is positive if and only if $q_m < q_0^{e^*}$. Thus, there exists at most two values of q_m such that equation (68) holds, and these values are independent of $\overline{U}$. In sum, generically we cannot have IC and IR binding with $q_M = q_m$.

3. If $q_M = \overline{q}$, then q_m must satisfy:

$$-\gamma\rho \int_0^{q_m} q \frac{\partial \phi}{\partial e}(q|e^*)dq + \rho \int_{q_m}^{\overline{q}} q \frac{\partial \phi}{\partial e}(q|e^*)dq = C'(e^*) \tag{69}$$

$$-\gamma\rho \int_0^{q_m} q\phi(q|e^*)dq + \rho \int_{q_m}^{\overline{q}} q\phi(q|e^*)dq - C(e^*) = \overline{U} \tag{70}$$

The LHS of the IR in equation (70) is strictly decreasing in q_m. Thus, there exists at most one value of q_m such that equation (70) holds, and this value is strictly decreasing in $\overline{U}$. The derivative of the LHS of IC in equation (69) with respect to q_m is $q_m \frac{\partial \phi}{\partial e}(q_m|e^*)\left(-\gamma\rho - \rho\right)$, which by MLRP and definition of $q_0^{e^*}$ is positive if and only if $q_m < q_0^{e^*}$. Thus, there exists at most two values of q_m such that equation (70) holds, and these values are independent of $\overline{U}$. In sum, generically we cannot have IC and IR binding with $q_M = \overline{q}$.

Case 2: $v'' < 0$.

We describe the optimal contract when the IC binds. By Lemma 1, when the IC in equation (2) is binding, it can be replaced by the FOC:

$$\int_0^{\overline{q}} u(w(q), q) \frac{\partial \phi}{\partial e}(q|e^*)dq = C'(e^*). \tag{71}$$

For now, ignore the monotonicity constraint. We will verify below that the optimal contract thus derived is monotonic.

Combining the agent's monotonicity constraint ($w(q)$ is nondecreasing in q) in equation (5) and Lemma 6, for some $q_m \in [0, \overline{q}]$ we have: $w(q) = 0$ for $q \in [0, q_m)$, and $w(q) \in [w^*(q), q]$ for

$q \geq q_m$ because of principal limited liability. This implies that $u(w,q) = v(w)$ for $q \geq q_m$. That is, for a given q_m, the relaxed optimization problem that gives the optimal contract to induce effort $e^* = e^T$ can be rewritten as:

$$\min_{w(q)} \int_{q_m}^{\overline{q}} w(q)\phi(q|e^*)dq \tag{72}$$

$$\text{s.t.} \quad \int_0^{q_m} u(0,q)\frac{\partial \phi}{\partial e}(q|e^*)dq + \int_{q_m}^{\overline{q}} v(w(q))\frac{\partial \phi}{\partial e}(q|e^*)dq = C'(e^*) \tag{73}$$

$$\int_0^{q_m} u(0,q)\phi(q|e^*)dq + \int_{q_m}^{\overline{q}} v(w(q))\phi(q|e^*)dq \geq \overline{U} \tag{74}$$

$$w(q) \in [w^*(q), q] \ \forall q \tag{75}$$

We henceforth consider the subset of values of q_m such that the optimization problem in equations (72)-(75) has a solution (the optimization problem has a solution for some q_m because of equations (6) and (15)). Using the notation in Jewitt, Kadan, and Swinkels (2008), we have $\underline{m}(q) = w^*(q)$ and $\overline{m}(q) = q$. We can apply Proposition 1 in their paper to derive the optimal contract on $[q_m, \overline{q}]$ given that the payment $w(q)$ is 0 on $[0, q_m)$ (note that the first terms on the LHS of equations (73) and (74) are independent of $w(q)$ and can therefore be treated as constants in the optimization problem in equations (72)-(75)). In sum, for some q_m, the optimal contract is defined implicitly by:

$$\frac{1}{u'_w(w(q),q)} = \begin{cases} \frac{1}{u'_w(0,q)} & \text{for } q \leq q_m \\ \frac{1}{v'(w^*(q))} & \text{for } q > q_m \text{ and } \lambda_{IR} + \lambda_{IC}LR(q|e^*) < \frac{1}{v'(w^*(q))} \\ \lambda_{IR} + \lambda_{IC}LR(q|e^*) & \text{for } q > q_m \text{ and } \frac{1}{v'(w^*(q))} < \lambda_{IR} + \lambda_{IC}LR(q|e^*) < \frac{1}{v'(q)} \\ \frac{1}{v'(q)} & \text{for } q > q_m \text{ and } \frac{1}{v'(q)} < \lambda_{IR} + \lambda_{IC}LR(q|e^*) \end{cases}$$

with $\lambda_{IR} \geq 0$ and $\lambda_{IC} > 0$, which are the Lagrange multipliers associated respectively with the constraints (74) and (73), and which therefore depend on q_m (in general, these are not the Lagrange multipliers associated with the IR and IC of the original optimization problem). Equivalently:

$$w(q) = \begin{cases} 0 & \text{for } q \leq q_m \\ w^*(q) & \text{for } q > q_m \text{ and } \lambda_{IR} + \lambda_{IC}LR(q|e^*) < \frac{1}{v'(w^*(q))} \\ v'^{-1}\left(1/\left(\lambda_{IR} + \lambda_{IC}LR(q|e^*)\right)\right) & \text{for } q > q_m \text{ and } \frac{1}{v'(w^*(q))} < \lambda_{IR} + \lambda_{IC}LR(q|e^*) < \frac{1}{v'(q)} \\ q & \text{for } q > q_m \text{ and } \frac{1}{v'(q)} < \lambda_{IR} + \lambda_{IC}LR(q|e^*) \end{cases}.$$

Proof of Corollary 1:

When the IC is binding, with a contract as in Proposition 2, the IC can be written as:

$$\int_0^{q_m} u(0,q)\frac{\partial \phi}{\partial e}(q|e^*)dq + \int_{q_m}^{\overline{q}} v(w(q))\frac{\partial \phi}{\partial e}(q|e^*)dq = C'(e^*) \tag{76}$$

In this equation, only the first term on the LHS depends on γ: $u(0,q) = v(w^*(q)) - (1+\gamma)w^*(q)$. Furthermore, with $q_m \le q_0^{e^T}$ and a binding IC which implies $e^* = e^T$, we have $\frac{\partial \phi}{\partial e}(q|e^*) < 0$ for any $q < q_m$.

With a contract as in Proposition 2, the first derivatives of the LHS of the IC and IR with respect to γ are respectively:

$$-\rho \int_0^{q_m} q \frac{\partial \phi}{\partial e}(q|e^*)dq > 0 \qquad \text{and} \qquad -\rho \int_0^{q_m} q\phi(q|e^*)dq < 0.$$

Thus, following a marginal change in γ, the contract must change in a way that decreases the LHS of the IC and increases the LHS of the IR.

With $v'' = 0$, $q_N = q_M$, so that the contract is fully described by q_m and q_M. With a contract as in Proposition 2, the first derivatives of the LHS of the IC with respect to q_m and q_M when $q_N = q_M$ are respectively:

$$-(1+\gamma)\rho q_m \frac{\partial \phi}{\partial e}(q_m|e^*) > 0 \qquad \text{and} \qquad (v(\rho q_M) - v(q_M))\frac{\partial \phi}{\partial e}(q_M|e^*) < 0.$$

With a contract as in Proposition 2, the first derivatives of the LHS of the IR in equation (65) with respect to q_m and q_M are respectively:

$$-(1+\gamma)\rho q_m \phi(q_m|e^*) < 0 \qquad \text{and} \qquad (v(\rho q_M) - v(q_M))\phi(q_M|e^*) < 0.$$

In sum, following a marginal increase in γ, an increase in both q_m and q_M would strictly decrease the LHS of the IR, while an increase in q_m and a decrease in q_M would strictly increase the LHS of the IC. Therefore, the only changes in q_m and q_M that leave the LHS of both the IC and IR unchanged overall following an increase in γ involve a decrease in q_m.

Online Appendix
A Theory of Fair CEO Pay

Pierre Chaigneau, Alex Edmans, Daniel Gottlieb

B Nonlinear Model

The agent's utility function is now:

$$u(w,q) \equiv \min\{v(w), \nu(w,q)\} \tag{77}$$

where $v(w)$ is utility over money alone, which is increasing and concave ($v' > 0$, $v'' \leq 0$), with $v(0) = 0$.[9] The term $\nu(w,q)$ is the agent's utility when his payment is below the fair wage, which in turn depends on output. We assume that $\nu(0,q) \leq 0$, $\nu_q(w,q) < 0$ (higher output raises the fair wage and thus lowers utility), $\nu_w(w,q) > 0$, $\lim_{w\searrow 0} \nu_w(w,q) > \lim_{w\searrow 0} v'(w)\ \forall\, q$ (the utility function is always steeper below the fair wage than above it), $\nu_{ww}(w,q) \geq 0$, $\nu_{wq} = 0$, and $\nu_{qq} = 0$.

For any given q, the functions $v(w)$ and $\nu(w,q)$ intersect on $(0,\infty)$ at most once.[10] Let this point, if it exists, be denoted by $w^*(q)$, i.e. $v(w^*(q)) \equiv \nu(w^*(q), q)$. At this point, $v'(w) < \nu_w(w,q)$, yielding a kink in $u(w,q)$ as a function of w at $w = w^*(q)$. Thus, $w^*(q)$ captures the agent's perceived fair wage, but it need no longer be linear in output. The utility function (77) exhibits not only loss aversion, but also concavity above the fair wage $w^*(q)$ and convexity below it, as in prospect theory.[11]

We also assume the following:

$$\overline{U} + C(e^*) > 0. \tag{78}$$

$$\int_0^{\overline{q}} v(w^*(q))\phi(q|e^*)dq - C(e^*) \geq \overline{U}, \quad \text{where } e^* \text{ satisfies (2) with } w(q) = w^*(q)\ \forall q, \tag{79}$$

Assumption (78) implies that a contract that always pays zero would violate the IR. Assumption (79) implies that a contract that always pays the fair wage satisfies the IR.

Propositions 4 and (5) study the case without moral hazard and the case with a binding IC, respectively.

Proposition 4 *(Zero target effort level, nonlinear model): Fix $e^T = 0$. If $\overline{U}$ is large enough, the principal implements $e^* > 0$. If $\overline{U}$ is small enough, the principal implements $e^* = 0$. In any case, the optimal contract is such that $w'(q) \geq 0$ with a strict inequality on a non-empty subinterval, and $w(q) \leq w^*(q)$ for any q.*

[9] This specification for the function $v(w)$ includes CRRA utility with relative risk aversion less than 1, and $v(w) = \ln(w+1)$.

[10] Indeed, for $w = 0$ and any q, we have $v(0) \geq \nu(0,q)$. In addition, for any q, $v(w)$ is weakly concave in w whereas $\nu(w,q)$ is weakly convex in w.

[11] However, the model does not exhibit probability weighting as in prospect theory.

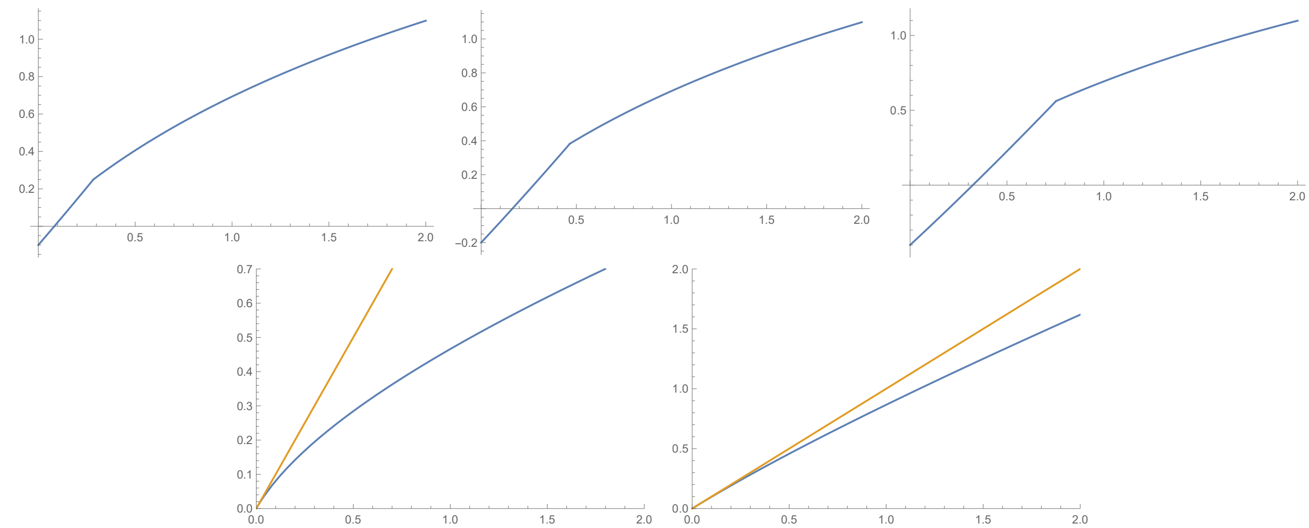

Figure 3: Top row: the function $u(w)$ defined in equation (77) as a function of w for $v(w) = \ln(w+1)$ and $\nu(w,q) = (w+1)^{1.2} - 1 - \frac{1}{5}q$ with $q = 0.5$ on the left, $q = 1$ in the middle, and $q = 2$ on the right. Bottom row: the blue line is the fair wage $w^*(q)$ defined by $v(w^*(q)) \equiv \nu(w^*(q), q)$ as a function of q for $v(w) = \ln(w+1)$ and $\nu(w,q) = (w+1)^{1.2} - 1 - \frac{1}{5}q$ on the left, $v(w) = \sqrt{w+1} - 1$ and $\nu(w,q) = w - \frac{1}{2}q$ on the right. The orange line is principal LL.

Proposition 5 *(Binding incentive constraint, nonlinear model) For the program (1)-(5) with a binding IC, the optimal contract is such that:*

$$w(q) = \begin{cases} 0 & \text{for } q \in [0, q_m) \\ w^*(q) & \text{for } q \in [q_m, q_M] \\ v'^{-1}\left(1/\left(\lambda_1 + \lambda_2 LR(q|e^*)\right)\right) & \text{for } q \in [q_M, q_N] \\ q & \text{for } q \in [q_N, \overline{q}] \end{cases}.$$

The optimal contract is given by four regions. As in the linear model, there are three regions in which the agent is paid zero, the fair wage, and the entire output. However, there is an additional region, given by $q \in [q_M, q_N]$, where output is sufficiently high that the principal pays more than the fair wage. It is inefficient to give the entire output, since the agent exhibits diminishing marginal utility and so does not value this additional reward highly. Thus, unlike in linear model, the optimal contract is continuous at q_M. As output rises above q_M, the likelihood ratio increases further and so the actual wage exceeds the fair wage by more. The contract will generally be convex between q_M and q_N.[12] For $q > q_N$, the likelihood ratio is so high that the principal pays the entire output.

If $e^T = 0$, i.e., the IC is slack, then the principal chooses the cheapest contract that satisfies the

[12]However, the contract will be concave if the likelihood ratio is concave, so that very high output is only slightly more indicative of effort, and if risk aversion is sufficiently important compared to prudence (see Chaigneau, Sahuguet and Sinclair-Desgagné, 2017). The latter condition will typically not be satisfied for CEOs who have low relative risk aversion due to their wealth.

IR; Proposition 5 shows that this involves a positive sensitivity of pay to performance, all the more that the agent's reservation utility $\overline{U}$ is high. Intuitively, the agent is never paid more than the fair wage, but he is paid at the fair wage for more outputs when $\overline{U}$ is higher. Thus, $\overline{U}$ is sufficiently high, the wage is increasing in output and fair for most outputs, so that it induces effort and the principal obtains effort "for free", as in Proposition 1.

This result is in stark contrast to the case without fairness concerns. In the standard model of Holmström (1979) with a risk-neutral principal and a risk-averse agent, eliciting higher effort is always more costly to the principal as it requires an output-contingent wage and thus inefficient risk-sharing.[13] As a result, any effort level in $\mathbb{R}_+$ can in principle be optimal, depending on model parameters. This is not true with fairness concerns. Providing low effort incentives either requires paying unfair wages for high outputs (which reduces expected utility and fails to satisfy the IR) or paying above the fair wage for low outputs (which is costly). Without fairness concerns, it is costly to incentivize high effort levels; with fairness concerns, it may be costly to incentivize low effort levels as doing so requires offering unfair pay. A by-product of fair pay is that it incentivizes effort, even if such incentives are unnecessary. This result may extend beyond the C-suite; for example, equity might be given to rank-and-file employees, despite their limited incentive effect, if they believe it is fair to share in the firm's fortunes.[14]

While paying the fair wage for a range of outputs helps satisfy the IR, doing so for all outputs would give the agent rents. The question then becomes: at which outputs does the firm pay below the fair wage, and how much below does it pay? With $\nu_{ww} > 0$, the agent's utility is non-concave below the fair wage. Thus, if the firm pays below the fair wage, it is efficient to pay him zero. Since the fair wage is increasing in output, the disutility from zero wages is also increasing in output, and so he should be paid zero for low output levels.

Example 2 illustrates how the outside option and the cost of effort affect the contract.

Example 2 *Consider $v(w) = \ln(w+1)$,[15] and an output that follows a truncated normal distribution on $(0, \infty)$ with e^* and $\sigma = 1$, $LR(q|e^*) \propto q +$ constant, and $v'^{-1}\left(1/\left(\lambda_1 + \lambda_2 LR(q|e^*)\right)\right)$ is linear in q. The contract is illustrated in Figure 4.*

In panel (a) of Figure 4, the IC is nonbinding (i.e. $e^* > e^T$), but the IR is binding, and so the principal pays the fair wage for low outputs. In panel (b), a lower $\overline{U}$ leads to a nonbinding IR and a binding IC (i.e. $e^* = e^T$). The contract is now driven by incentive considerations and so the fair wage is no longer paid for outputs with a negative likelihood ratio.[16] In panel (c), the cost of effort is higher than in panel (b), requiring the principal to increase incentives. She does so by paying the fair wage rather than zero for a larger subset of outputs with a positive likelihood ratio,

[13] If the agent is risk-neutral and protected by limited liability, implementing higher effort requires the firm to offer him a higher payment upon success and thus a higher expected wage.

[14] Oyer (2004) offers a different explanation for this practice – that equity-based pay ensures that wages constantly match outside opportunities.

[15] This yields $v'^{-1}(w) = \frac{1}{w} - 1$, so that $v'^{-1}\left(1/\lambda LR(q|e^*)\right) = \lambda LR(q|e^*) - 1$.

[16] When the IC is binding so that $e^* = e^T$, the likelihood ratio is positive for $LR(q|e^T) > q_0^{e^T} \approx 5$. The approximation is due to the use of the truncated normal distribution with $e = 5$ and $\sigma = 1$.

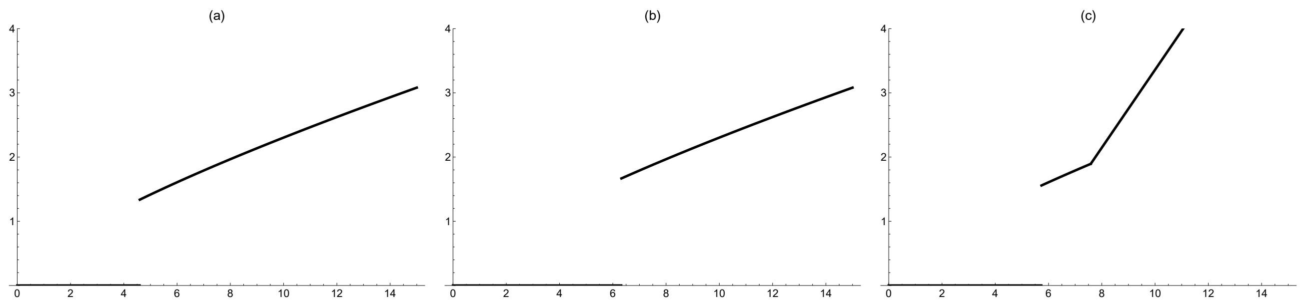

Figure 4: The contract $w(q)$ as a function of q. The agent's preferences are as in Example 2 with $\nu(w,q) = (w+1)^{1.2} - 1 - \frac{1}{5}q$, $C(e) = \frac{c}{10}\left(\frac{e}{e^T}\right)^{10}$, and $e^T = 5$. (a): $\overline{U} = 0$ and $c = 0.02$. (b): $\overline{U} = -2$ and $c = 0.02$. (c): $\overline{U} = -2$ and $c = 0.05$.

and above the fair wage for very high outputs.

Proof of Proposition 4:

In the optimization problem with a nonbinding IC, the IR for $e^* \geq 0$ must be binding. Suppose that it is not. Then, the contract that solves the optimization problem in equations (1), (4), and (5) is simply $w(q) = 0$ for any q, so that, using equation (78) at any effort e^* with $\nu(0,q) \leq 0$ by assumption:

$$\int_0^{\overline{q}} u(0,q)\phi(q|e^*)dq = \int_0^{\overline{q}} \nu(0,q)\phi(q|e^*)dq \leq 0$$

i.e. IR is not satisfied because of equation (78), a contradiction.

The relaxed optimization problem with a nonbinding IC, a binding IR, and the FOA, is:

$$\min_{w(q)} \int_0^{\overline{q}} w(q)\phi(q|e^*)dq \tag{80}$$

$$\text{s.t.} \quad \int_0^{\overline{q}} u(w(q),q)\phi(q|e^*)dq - C(e^*) = \overline{U} \tag{81}$$

$$\int_0^{\overline{q}} u(w(q),q)\frac{\partial\phi}{\partial e}(q|e^*)dq = C'(e^*) \tag{82}$$

$$0 \leq w(q) \leq q \tag{83}$$

$$w(q) \geq w(q') \; \forall q > q' \tag{84}$$

Lemma 8 *If the optimal contract induces $e^* = 0$, it is such that $w(q) \leq w^*(q)$ on any non-empty subinterval of $[0, \overline{q}]$.*

Proof. We will show that a contract such that $w(q) > w^*(q)$ for some q is dominated. Since the contract must involve $w(q) > 0$ for some q as shown above and $w^*(0) = 0$, this also implies that the optimal contract is such that $w'(q) > 0$ for some q because of equation (84).

Denote by $w^A(q)$ a contract that satisfies the constraints on contracting and the participation constraint. A contract such that $w(q) \geq w^*(q)$ for all q with a strict inequality for some nonempty subinterval(s) is dominated. Indeed, with the assumption in equation (79), there exists a cheaper contract which still satisfies the constraints on contracting and the participation constraint. Now consider a contract $w^A(q)$ such that $w^A(q) > w^*(q)$ on a nonempty subinterval of $[0, \overline{q}]$ denoted by $\mathcal{Q}_a$, $w^A(q) < w^*(q)$ on a nonempty subinterval of $[0, \overline{q}]$ denoted by $\mathcal{Q}_b$, and $w^A(q) = w^*(q)$ on a possibly empty subinterval of $[0, \overline{q}]$ denoted by $\mathcal{Q}_c$.

For $q \in \mathcal{Q}_a$, since $v'' < 0$:

$$v(w^A(q)) - v(w^*(q)) < v'(w^*(q)) \left(w^A(q) - w^*(q)\right) < v'(0) \left(w^A(q) - w^*(q)\right) \tag{85}$$

For $q \in \mathcal{Q}_b$, since $\nu'' \geq 0$ and because of the assumption $\lim_{w \searrow 0} \nu_w(w, q) > \lim_{w \searrow 0} v'(w) \; \forall \, q$:

$$\nu(w^*(q), q) - \nu(w^A(q), q) \geq \nu'_w(w^*(q), q) \left(w^*(q) - w^A(q)\right) > v'(0) \left(w^*(q) - w^A(q)\right) \tag{86}$$

Since the contract $w^A(q)$ satisfies the participation constraint:

$$\int_{\mathcal{Q}_a} v(w^A(q)) f(q|e^*) dq + \int_{\mathcal{Q}_b} \nu(w^A(q), q) f(q|e^*) dq + \int_{\mathcal{Q}_c} v(w^*(q)) f(q|e^*) dq - C(e^*) = \overline{U}_0, \tag{87}$$

where $\overline{U}_0 \geq \overline{U}$. Denote by $\mathcal{Q}_{b1}$ and $\mathcal{Q}_{b2}$ two subintervals of $\mathcal{Q}_b$ such that $\mathcal{Q}_{b1} \cup \mathcal{Q}_{b2} = \mathcal{Q}_b$ and $q_2 < q_1$ for any $q_2 \in \mathcal{Q}_{b2}$ and $q_1 \in \mathcal{Q}_{b1}$. Because of the assumption in equation (79), there exists a contract with payment $w^*(q)$ for any q except on the subinterval $\mathcal{Q}_{b2}$, where payment is $w^A(q)$ such that $w^A(q) < w^*(q)$, that induces effort $\hat{e}$, such that:

$$\begin{aligned}\int_{\mathcal{Q}_a} v(w^*(q)) f(q|e^*) dq + \int_{\mathcal{Q}_{b1}} \nu(w^*(q), q) f(q|e^*) dq + \int_{\mathcal{Q}_{b2}} \nu(w^A(q), q) f(q|e^*) dq \\ + \int_{\mathcal{Q}_c} v(w^*(q)) f(q|e^*) dq - C(e^*) = \overline{U}_0\end{aligned} \tag{88}$$

$$\begin{aligned}\int_{\mathcal{Q}_a} v(w^*(q)) f(q|\hat{e}) dq + \int_{\mathcal{Q}_{b1}} \nu(w^*(q), q) f(q|\hat{e}) dq + \int_{\mathcal{Q}_{b2}} \nu(w^A(q), q) f(q|\hat{e}) dq \\ + \int_{\mathcal{Q}_c} v(w^*(q)) f(q|\hat{e}) dq - C(\hat{e}) \geq \overline{U}_0\end{aligned} \tag{89}$$

Equation (89) is satisfied as an equality under effort $\hat{e} = e^*$. Since by construction the new contract provides the agent with the same expected utility net of effort cost $\overline{U}_0$ under effort e^* associated with the initial contract, if the agent is better off exerting a different effort level $\hat{e}$, then equation (89) is satisfied as an inequality. By construction, we have:

$$\begin{aligned}\int_{\mathcal{Q}_a} \left(v(w^A(q)) - v(w^*(q))\right) f(q|e^*) dq + \int_{\mathcal{Q}_{b1}} \left(\nu(w^A(q), q) - \nu(w^*(q), q)\right) f(q|e^*) dq \\ + \int_{\mathcal{Q}_{b2}} \left(\nu(w^A(q), q) - \nu(w^A(q), q)\right) f(q|e^*) dq + \int_{\mathcal{Q}_c} \left(v(w^*(q)) - v(w^*(q))\right) f(q|e^*) dq = 0\end{aligned} \tag{90}$$

Applying equations (85) and (86) to subintervals $\mathcal{Q}_a$ and $\mathcal{Q}_{b1}$, respectively:

$$\begin{aligned}
& \int_{\mathcal{Q}_a} v(w^A(q))f(q|e^*)dq - \int_{\mathcal{Q}_a} v(w^*(q))f(q|\hat{e})dq + \int_{\mathcal{Q}_{b1}} \nu(w^A(q),q)f(q|e^*)dq - \int_{\mathcal{Q}_{b1}} \nu(w^*(q),q)f(q|\hat{e})dq \\
& + \int_{\mathcal{Q}_{b2}} \nu(w^A(q),q)f(q|e^*)dq - \int_{\mathcal{Q}_{b2}} \nu(w^A(q),q)f(q|\hat{e})dq + \int_{\mathcal{Q}_c} v(w^*(q))f(q|e^*)dq - \int_{\mathcal{Q}_c} v(w^*(q))f(q|\hat{e})dq \\
< \; & \int_{\mathcal{Q}_a} \left(v(w^*(q)) + v'(0)\left(w^A(q) - w^*(q)\right)\right) f(q|e^*)dq - \int_{\mathcal{Q}_a} v(w^*(q))f(q|\hat{e})dq \\
& + \int_{\mathcal{Q}_{b1}} \left(v(w^*(q)) + v'(0)\left(w^A(q) - w^*(q)\right)\right) f(q|e^*)dq - \int_{\mathcal{Q}_{b1}} \nu(w^*(q),q)f(q|\hat{e})dq \\
& + \int_{\mathcal{Q}_{b2}} \nu(w^A(q),q)f(q|e^*)dq - \int_{\mathcal{Q}_{b2}} \nu(w^A(q),q)f(q|\hat{e})dq + \int_{\mathcal{Q}_c} v(w^*(q))f(q|e^*)dq - \int_{\mathcal{Q}_c} v(w^*(q))f(q|\hat{e})dq \\
\leq \; & \int_{\mathcal{Q}_a} v'(0)\left(w^A(q) - w^*(q)\right) f(q|e^*)dq + \int_{\mathcal{Q}_{b1}} v'(0)\left(w^A(q) - w^*(q)\right) f(q|e^*)dq \\
= \; & v'(0) \int_{\mathcal{Q}_a \cup \mathcal{Q}_{b1}} \left(w^A(q) - w^*(q)\right) f(q|e^*)dq \\
= \; & v'(0) \int_0^{\overline{q}} \left(w^A(q) - w^*(q)\right) f(q|e^*)dq
\end{aligned} \tag{91}$$

where we used equations (88) and (89) to get the second inequality, and the fact that both contracts have the same payment on subintervals $\mathcal{Q}_{b2}$ and $\mathcal{Q}_c$ to get the final equality. Using equation (90) and $v' > 0$, the last line of equation (91) is greater than zero, which means that the expected cost of the new contract is higher under effort e^*.

There are two cases. In any case, we will establish that a contract such that $w(q) > w^*(q)$ for some q is dominated.

First, suppose that the effort induced by the new contract described above is lower than (or equal to) the level of effort e^* associated with the initial contract $w^A(q)$. Since compensation is increasing in q, this weakly reduces the expected cost of the contract, which shows that the initial contract is dominated.

Second, suppose that the level of effort induced by the new contract described above is strictly higher than the level of effort e^* associated with the initial contract $w^A(q)$. Denote by q_a the highest level of q such that, under the new contract, $w(q) = w^*(q)$. By construction of this new contract, we have $w(q) = w^*(q)$ for $q \leq q_a$, and $w(q) \in [w(q_a), w^*(q))$ for $q \geq q_a$. Now suppose the payment under this new contract is set to $w(q_b)$ for any $q \geq q_b$, where q_b is set so that the expected utility of the contract is equal to $\overline{U}$. The threshold q_b is strictly increasing in $\overline{U}$:

$$\int_0^{q_b} v(w^*(q))\phi(q|e^*)dq + \int_{q_b}^{\overline{q}} \nu(w(q_b),q)\phi(q|e^*)dq = \overline{U} \tag{92}$$

Thus, when $\overline{U}$ is sufficiently low, $q_b \leq q_a$, and the expected cost of the contract is reduced by the aforementioned perturbation. Moreover, when $\overline{U}$ is sufficiently low, the resulting contract induces

$e^* = 0$ since:

$$\int_0^{q_b} v(w^*(q))\frac{\partial \phi}{\partial e}(q|0)dq + \int_{q_b}^{\overline{q}} \nu(w(q_b), q)\frac{\partial \phi}{\partial e}(q|0)dq < 0. \tag{93}$$

This also establishes that the initial contract is dominated.

Now consider the case when $\overline{U}$ is sufficiently large: it is equal to the LHS of equation (79). Then the optimal contract pays $w^*(q)$ on any non-empty subinterval by the arguments above and equation (79). In this case, effort is strictly positive ($e^* > 0$) since for any e^*:

$$\int_0^{\overline{q}} v(w^*(q))\frac{\partial \phi}{\partial e}(q|e^*)dq > 0. \tag{94}$$

■

Proof of Proposition 5:

We describe the optimal contract when the IC binds. By Lemma 1, when the IC in equation (2) is binding, it can be replaced by the FOC:

$$\int_0^{\overline{q}} u(w(q), q)\frac{\partial \phi}{\partial e}(q|e^*)dq = C'(e^*). \tag{95}$$

For now, ignore the monotonicity constraint in equation (84). We will verify below that the optimal contract thus derived satisfies monotonicity. This part of the proof has two steps.

Lemma 9 *On any non-empty subinterval of* $[0, \overline{q}]$, *we have* $w(q) \notin (0, w^*(q))$.

Proof. Let $\breve{u}(q) := u(w(q), q)$, and let $u_w^{-1}(\cdot)$ be such that $u_w^{-1}(\breve{u}(q)) = w(q) \;\Leftrightarrow\; u_w^{-1}(u(w(q), q)) = w(q)$. The Lagrangian for the optimization problem in equations (80)-(83) is:

$$\begin{aligned}\mathcal{L} \;&=\; \int_0^{\overline{q}} u_w^{-1}(\breve{u}(q))\,\phi(q|e^*)dq - \lambda\left(\int_0^{\overline{q}} \breve{u}(q)\phi(q|e^*)dq - C(e^*) - \overline{U}\right)\\ &\quad -\mu\left(\int_0^{\overline{q}} \breve{u}(q)\frac{\partial \phi}{\partial e}(q|e^*)dq - C'(e^*)\right) - \lambda_{\text{LLA}}(q)u_w^{-1}(\breve{u}(q)) - \lambda_{\text{LLP}}(q)\left(q - u_w^{-1}(\breve{u}(q))\right)\end{aligned}$$

Note that the constraints are linear in $\breve{u}(q)$. The first-order necessary condition (FONC) with respect to $\breve{u}(q)$ at any given output q is:

$$\begin{aligned}&u_w^{-1\prime}(\breve{u}(q))\,\phi(q|e^*) - \lambda\phi(q|e^*) - \mu\frac{\partial \phi}{\partial e}(q|e^*) - \lambda_{\text{LLA}}(q)u_w^{-1\prime}(\breve{u}(q)) + \lambda_{\text{LLP}}(q)u_w^{-1\prime}(\breve{u}(q)) = 0\\ \Leftrightarrow\quad &u_w^{-1\prime}(\breve{u}(q)) = \lambda + \mu\frac{\frac{\partial \phi}{\partial e}(q|e^*)}{\phi(q|e^*)} + \frac{\lambda_{\text{LLA}}(q)}{\phi(q|e^*)}u_w^{-1\prime}(\breve{u}(q)) - \frac{\lambda_{\text{LLP}}(q)}{\phi(q|e^*)}u_w^{-1\prime}(\breve{u}(q))\end{aligned} \tag{96}$$

where $\lambda_{\text{LLA}}(q) = 0$ for $w(q) > 0 \;\Leftrightarrow\; u_q > u(0, q)$, and $\lambda_{\text{LLP}}(q) = 0$ for $w(q) < q \;\Leftrightarrow\; u_q < u(q, q)$.

The proof is by contradiction. For a given q, suppose that $w(q) \in (0, w^*(q)) \;\Leftrightarrow\; u_q \in (u(0, q), u(w^*(q), q))$, which implies $\lambda_{\text{LLA}}(q) = 0$ and $\lambda_{\text{LLP}}(q) = 0$ by definition of these Lagrange multipliers, and also implies $u(w(q), q) = \nu(w(q), q)$ by definition of the utility function. The function $\nu(w, q)$ is increasing and convex in w. This implies that u_w^{-1} is increasing and concave. Thus, the FONC does not describe an optimum to the minimization problem, and furthermore the optimal $w(q)$ is not in the interval $(0, w^*(q)) \Leftrightarrow$ the optimal u_q is not in the interval $(u(0, q), u(w^*(q), q))$. ■

Combining the agent's monotonicity constraint ($w(q)$ is nondecreasing in q) in equation (5) and Lemma 9, for some $q_m \in [0, \overline{q}]$ we have: $w(q) = 0$ for $q \in [0, q_m)$, and $w(q) \in [w^*(q), q]$ for $q \geq q_m$ because of principal limited liability. This implies that $u(w, q) = v(w)$ for $q \geq q_m$. That is, for a given q_m, the relaxed optimization problem that gives the optimal contract to induce effort $e^* = e^T$ can be rewritten as:

$$\min_{w(q)} \int_{q_m}^{\overline{q}} w(q)\phi(q|e^*)dq \tag{97}$$

$$\text{s.t.} \quad \int_0^{q_m} u(0, q)\frac{\partial \phi}{\partial e}(q|e^*)dq + \int_{q_m}^{\overline{q}} v(w(q))\frac{\partial \phi}{\partial e}(q|e^*)dq = C'(e^*) \tag{98}$$

$$\int_0^{q_m} u(0, q)\phi(q|e^*)dq + \int_{q_m}^{\overline{q}} v(w(q))\phi(q|e^*)dq \geq \overline{U} \tag{99}$$

$$w(q) \in [w^*(q), q] \; \forall q \tag{100}$$

We henceforth consider the subset of values of q_m such that the optimization problem in equations (97)-(100) has a solution (the optimization problem has a solution for some q_m because of equations (6) and (79)). Using the notation in Jewitt, Kadan, and Swinkels (2008), we have $\underline{m}(q) = w^*(q)$ and $\overline{m}(q) = q$. We can apply Proposition 1 in their paper to derive the optimal contract on $[q_m, \overline{q}]$ given that the payment $w(q)$ is 0 on $[0, q_m)$ (note that the first terms on the LHS of equations (98) and (99) are independent of $w(q)$ and can therefore be treated as constants in the optimization problem in equations (97)-(100)). In sum, the optimal contract is defined implicitly by:

$$\frac{1}{u'_w(w(q), q)} = \begin{cases} \frac{1}{u'_w(0,q)} & \text{for } q \leq q_m \\ \frac{1}{v'(w^*(q))} & \text{for } q > q_m \text{ and } \lambda_{IR} + \lambda_{IC} LR(q|e^*) < \frac{1}{v'(w^*(q))} \\ \lambda_{IR} + \lambda_{IC} LR(q|e^*) & \text{for } q > q_m \text{ and } \frac{1}{v'(w^*(q))} < \lambda_{IR} + \lambda_{IC} LR(q|e^*) < \frac{1}{v'(q)} \\ \frac{1}{v'(q)} & \text{for } q > q_m \text{ and } \frac{1}{v'(q)} < \lambda_{IR} + \lambda_{IC} LR(q|e^*) \end{cases}$$

with $\lambda_{IR} \geq 0$ and $\lambda_{IC} > 0$, which are the Lagrange multipliers associated respectively with the constraints (99) and (98), and which therefore depend on q_m (in general, these are not the Lagrange multipliers associated with the IR and IC of the original optimization problem). Equivalently:

$$w(q) = \begin{cases} 0 & \text{for } q \leq q_m \\ w^*(q) & \text{for } q > q_m \text{ and } \lambda_{IR} + \lambda_{IC} LR(q|e^*) < \frac{1}{v'(w^*(q))} \\ v'^{-1}\left(1/\left(\lambda_{IR} + \lambda_{IC} LR(q|e^*)\right)\right) & \text{for } q > q_m \text{ and } \frac{1}{v'(w^*(q))} < \lambda_{IR} + \lambda_{IC} LR(q|e^*) < \frac{1}{v'(q)} \\ q & \text{for } q > q_m \text{ and } \frac{1}{v'(q)} < \lambda_{IR} + \lambda_{IC} LR(q|e^*) \end{cases}.$$