A Theory of Fair CEO Pay*

Pierre Chaigneau
Queen's University

Alex Edmans
LBS, CEPR, and ECGI

Daniel Gottlieb
LSE

March 15, 2022

Abstract

This paper studies optimal executive pay when the CEO is concerned about fairness – if his wage falls below a perceived fair share of output, the CEO suffers a disutility loss that is increasing in the discrepancy. Fairness concerns do not lead to fair wages always being paid. To induce effort, the firm threatens the CEO with unfair wages if output is sufficiently low. The optimal contract resembles 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 continues to involve pay-for-performance, to address the CEO's fairness concerns and ensure his participation. Thus, the firm can implement strictly positive levels of effort "for free." This rationalizes pay-for-performance even if the CEO is intrinsically motivated and does not need effort incentives.

Keywords: Executive compensation, contract theory, moral hazard, fairness, loss aversion, prospect theory, reference points.
JEL Classification: D86, G32, G34, J33.

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

Standard moral hazard models of CEO compensation assume that the CEO cares about pay only for the consumption utility it provides. As a result, the marginal consumption utility of the additional pay from improving performance must be at least as great as the marginal cost of effort required to do. Such models have contributed substantially to our understanding of CEO pay, and generated many predictions that are consistent with observed contracts.

However, some features of real-life contracts are difficult to explain with standard CEO compensation theories. These models predict that pay should be smoothly increasing in output, but actual contracts feature discontinuities such as performance shares or bonuses that are forfeited if performance falls below a certain threshold. In addition to their predictions, the assumptions of standard models may not be consistent with the realities of the CEO setting. Many models of CEO pay are standard principal-agent models applied to the CEO setting. However, one important difference between CEOs and other agents is that CEOs are typically wealthy, and nearly all of their consumption needs are already met. Thus, it is not clear that consumption utility is the only, or even the most important, driver, of observed contracts.

Edmans, Gosling, and Jenter (2022) survey directors and investors on how they set and influence pay contracts. Both sets of respondents highlight how pay is not driven purely by the desire to provide consumption incentives, but significantly influenced by concerns for fairness. They also suggest that an important reference point, that directors, investors, and the CEO use to determine whether pay is fair, is firm value. If firm value has increased due to CEO effort, these parties believe that it is fair to reward the CEO for this increase. If firm value has increased/decreased due to luck outside the CEO's control, they believe that the CEO should share in this good/bad luck. The reference point of a share of firm value is consistent with the widely-replicated ultimatum game. If one party has been gifted an endowment, the other party believes that it is fair to be offered a sizable share, and will sacrifice his own consumption to punish an unfair offer even if it is strictly positive.

This paper studies optimal CEO pay when the CEO is motivated by both traditional consumption utility and fairness concerns. We model fairness concerns by specifying a reference point; the CEO suffers disutility if his wage falls below the reference point, the magnitude of which is increasing in the discrepancy. This reference point is increasing in the firm's output, which depends on both CEO effort and luck.

It may seem that fairness concerns should lead to the CEO always receiving a fair wage, but this turns out not to be the case. We start with a risk-neutral model that demonstrates the effect of fairness concerns in the most transparent possible setting. The fair wage is linear in output, i.e. the CEO believes that it is fair for him to receive a given share of output. If the actual wage is at least the fair wage, his utility equals the wage, as with standard risk neutrality. If the actual wage is below the fair wage, he suffers disutility which is linear in the discrepancy. The CEO's utility function is thus piecewise linear, with a slope of 1 above the fair wage and a slope exceeding 1 below it. The principal is also risk-neutral, and her goal is to find the cheapest contract to induce a given effort level out of a continuum. Both parties are protected by limited liability.

We show that the optimal contract involves a threshold below which the CEO is paid zero, and above which the CEO receives the fair wage, i.e. a constant share of output. This contradicts the intuition that the fairness concerns will lead to the CEO being paid fair wages for all output levels. Instead, fairness concerns mean that *un*fairness can be a powerful motivator. If output is sufficiently low that it is unlikely that the CEO has worked, the firm pays him the most unfair possible wage of zero. Only if output exceeds a lower threshold is the CEO paid the fair wage. Depending on parameter values, there may also be an additional upper threshold above which the CEO is paid the firm's entire output.

Innes (1990) showed that, with standard risk neutrality, the optimal contract is "live-or-die" – the agent receives zero if output is below a threshold, and the entire output above it. The intuition is that it is optimal to concentrate payments in the highest likelihood ratio states, i.e. pay the highest possible amount for sufficiently high outputs. However, 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 suffers significant disutility due to unfairness, which erodes his incentives to work. Thus, it is efficient to offer him a fair wage for intermediate output levels. Put differently, since the agent's utility function is steeper below than above the fair wage, it is efficient to lower the wage from the entire output to the fair wage for some high output levels, and use these savings to increase the wage from zero to the fair wage for some moderate output levels. We show that, as long as the CEO is not paid zero for outputs which are good news about effort, the threshold level of output above which the CEO is paid fair wages is decreasing in the CEO's fairness concerns. In addition, the range of outputs over which the CEO is paid fair wages is increasing the volatility of the output distribution. Intuitively, the more volatile the output distribution, the likelier it is that output will be moderate even if the CEO works, and so the more important it is to reward him with a fair wage rather than zero.

The contract resembles performance shares, which are frequently offered in reality (see the survey of Edmans, Gabaix, and Jenter (2017)). Standard models, such as Holmström (1979), do not predict discontinuous contracts. Innes (1990) predicts a sharp discontinuity where the CEO's pay increases from zero to the entire output, once output crosses a threshold, but such sharp discontinuities do not exist in reality. To obtain more realistic contracts, Innes (1990) assumes that the principal's payoff cannot be decreasing in output, otherwise the principal would exercise her control rights to "burn" output, or the agent would secretly inject his own funds into the company to inflate output. Innes's theory can either be interpreted as 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 would be illegal. Similarly, it would likely be illegal for the CEO to inject his own funds into the company to manipulate the stock price. Indeed, the prevalence of discontinuities in real-life executive compensation contracts suggests that these two justifications are not relevant. Our paper obtains realistic contracts without a monotonicity assumption to rule out discontinuities, and indeed the optimal contract involves a discontinuity. Performance shares provide fair wages if performance is good and unfair wages if performance is bad, to motivate good performance.

We then extend the model to the case of risk aversion. Now, the utility function takes a more general form – if the wage is fair, utility is increasing and concave in the wage; if the wage is unfair, it is a general function of both the wage and output which is increasing and convex in the former (as in prospect theory) and decreasing in the latter. In particular, the fair wage need no longer be a constant share of output, and utility loss from unfair wages need not be linear in the discrepancy. Despite this additional generality, we show that the basic features of the risk-neutral contract 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 very strong, the principal wishes to reward the CEO with more than the fair wage. However, since the CEO is risk-averse, it is inefficient to pay him the entire output.

We show that pay is increasing in output even when the incentive constraint is slack. When participation is the only constraint, it may seem that the most efficient way to satisfy this constraint is to pay the CEO a fair wage for all outputs – the prior argument that zero wages are needed to punish low effort no longer applies. However, guaranteeing the CEO a fair wage for all outputs would lead to the participation constraint being slack. To avoid giving the CEO rents, the firm pays him an unfair wage for some output levels. Since the CEO's utility function is convex below the fair wage, if the firm reduces the payment below the fair wage, it is optimal to reduce it all the way to zero. Thus, the firm pays the CEO zero for some output levels, rather than a moderately unfair wage for a greater range of output levels.

That pay is increasing in output even without an incentive constraint means that the firm can induce CEO effort "for free". In a standard moral hazard model, implementing higher effort is always costly to the firm.[1] In our model, since pay is optimally increasing in output to satisfy the participation constraint, lower effort levels will more costly to implement than certain higher effort levels, so they will never be induced. A frequent criticism of performance-related pay for CEOs is that it should not be necessary – the CEO should be intrinsically motivated to exert effort, and/or the board should monitor the CEO to ensure effort. Our model demonstrates that performance-related pay may be optimal not to induce effort, but to secure the participation of a CEO with fairness concerns.

This paper is related to the theoretical literature on executive compensation, recently surveyed by Edmans and Gabaix (2016) and Edmans, Gabaix, and Jenter (2017). While a small number of

[1] If the CEO is risk-neutral and protected by limited liability, this requires the firm to offer him a higher payment upon success and thus a higher expected wage; if the CEO is risk-averse, this requires the firm to offer him a more sensitive contract and thus a risk premium.

models focus on adverse selection or retention, the vast majority of these theories feature moral hazard, where pay only matters to the CEO by providing consumption utility. 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 CEO's utility is steeper below the fair wage than above it, as in loss aversion. The key innovation in our model is that the reference point depends on output.

An important literature has studied the effect of fairness concerns in contracts outside of the CEO setting. Fehr and Schmidt (1999) study optimal contracts in the presence of inequity aversion, where an agent dislikes another agent getting less than him, and dislikes even more another agent getting more than him. Sobel (2005) provides a survey of this literature.[2] In such models, the agents are all paid in the same units, and so it makes sense for agents to compare their consumption. However, these models do not apply to a CEO setting, where the firm's objective function is shareholder value, which is orders of magnitude in excess of CEO pay, and thus in different units. An inequity aversion explanation for performance-sensitive CEO pay is that shareholders feel sorry for a CEO who is not given a share of greater firm value, which seems at odds with common perceptions that fairer CEO pay would involve lower pay. In our model, it is the CEO who has non-standard preferences, rather than shareholders. Moreover, the CEO is only concerned for his own utility, unlike in social preference models where agents are concerned with other agents' utility.

1 The Model

We consider a standard contracting model with a principal (firm, "she") and an agent (manager, "he"). At time $t = -1$, the principal offers a contract to the agent. At $t = 0$, if the agent has accepted the contract, he chooses an effort level $e \in [0, \bar{e}]$, where $\bar{e}$ can be finite or infinite. The agent's cost of exerting effort e is $C(e)$, where $C(\cdot)$ is continuous and non-decreasing. We have $C'(e) > 0$ for $e > 0$ and $C'(0) = 0$. As is standard, effort can refer not only to working rather than shirking, but also to choosing projects to maximize firm value rather than private benefits or not diverting cash flows. At $t = 1$, output $q \in [0, \overline{q}]$ is realized and the principal is paid a wage $w(q)$ that depends on output. Both the principal and agent are protected by limited liability, so that $0 \leq w(q) \leq q \; \forall q$. The agent's utility function is $u(w, q)$, which depends on the actual wage w, and may depend on output q if it affects the wage that he believes to be fair. The agent's reservation utility is given by $\overline{U}$.

The principal does not observe the agent's effort, but observes the realization of output. Output is continuously distributed according to a probability density function ("PDF") $\phi(q|e)$ that satisfies the monotone likelihood ratio property ("MLRP") and is continuously differentiable in q and in e

[2] In Rabin (1993), an agent may put positive or negative weight on the other agent's utility, depending on his assessment of her intentions. Empirically, Fehr, Klein, and Schmidt (2007) show how inequity aversion leads to a principal paying an agent a discretionary bonus upon good performance, even though such a bonus is unenforceable. Fehr, Kirchsteiger, and Riedl (1993) show that workers reciprocate a fair wage with higher effort in the next period, even though this higher effort is unenforceable. Charness and Rabin (2002) use experiments to distinguish between various models of social preferences.

with a continuous cross derivative: $\frac{\partial^2}{\partial q \partial e}\phi(q|e)$ is continuous in q.

As is standard, the principal's problem is to choose a wage schedule $w(q)$ to maximize firm value net of pay, $q - w(q)$, to induce an effort level of at least e^T. Her program is given as follows:

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

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

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

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

where (2) is the incentive compatibility constraint ("IC") and (3) is the individual rationality constraint ("IR").

The above formulation captures fairness concerns in the simplest possible way. We use the standard moral hazard model with continuous effort and continuous output, with the only departure being the specification of the agent's utility function. As a result, any deviation in the optimal contract from standard moral hazard models can be attributed to the utility function. Another formulation, which would follow the ultimatum game more literally, would be to have a multi-period model where the agent responds to his first-period pay by choosing effort in the second period. Then, if he is offered unfair pay, he may withhold effort and destroy total surplus, similar to the respondent in the ultimatum game refusing the proposed share and leading to both parties receiving zero. Our one-period model already captures the agent's preferences for fairness, without having to introduce multiple periods, since he knows the contract at the time he chooses his effort level, and will withhold effort in the current period if he anticipates that he will not be paid fairly for his effort. This allows our results to be compared with standard one-period moral hazard models without fairness concerns, such as Holmström (1979) and Innes (1990).

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

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

and let q_0^e be the output such that the likelihood ratio is zero for an effort of e: $LR(q_0^e|e) \equiv 0$.

To guarantee that an optimal contract exists, we assume:

$$\int_0^{q_0^{e^*}} u(0, q)\frac{\partial}{\partial e}\phi(q|e^*)dq + \int_{q_0^{e^*}}^{\overline{q}} u(q, q)\frac{\partial}{\partial e}\phi(q|e^*)dq \geq C'(e^*)$$

The above inequality means that 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 be sufficient to induce effort e^*. If it is not satisfied, then no contract that satisfies bilateral limited liability can implement effort e^*.

In the analysis that follows, we will distinguish between cases in which the participation constraint is binding or nonbinding, as this will affect the optimal contract. The participation constraint is nonbinding if:[3]

$$\overline{U} + C(e^*) \leq \int_0^{\overline{q}} u(0,q)\phi(q|e^*)dq.$$

The first-order approach ("FOA") simplifies the analysis of moral hazard models with a continuum of efforts by allowing a continuum of incentive constraints (equation (2)) to be replaced by the local incentive constraint that prevents local deviations. Chaigneau, Edmans, and Gottlieb (2022) note that standard conditions used to justify the FOA with commonly-used output distributions cannot be used in models with limited liability, and derive a new sufficient condition for the validity of the FOA under limited liability. Lemma 1 below derives a similar condition in the setting considered in this paper, in which the utility function also depends on output. Let K_e^+ and K_e^- denote the integral of the positive and negative parts of the second derivative of the joint distribution $\phi(q|e)$ with respect to effort:

$$K_e^+ := \int_0^{\overline{q}} \max\left\{\frac{\partial^2}{\partial e^2}\phi(q|e),\ 0\right\} dq, \tag{5}$$

$$K_e^- := \int_0^{\overline{q}} \min\left\{\frac{\partial^2}{\partial e^2}\phi(q|e),\ 0\right\} dq. \tag{6}$$

Lemma 1 *(First-Order Approach): Suppose that*

$$\int_0^{\overline{q}} \left(K_e^- u(0,q) + K_e^+ u(q,q)\right) dq < C''(e) \tag{7}$$

for all $e \in (0, \bar{e})$. Then, the FOA is valid.

We henceforth assume that this condition holds for the relevant specification of the utility function.

We start in Section 2 by considering a risk-neutral agent with fairness concerns. This demonstrates the effect of fairness concerns in the simplest possible setting, where the utility function is piecewise linear. Section 3 incorporates the effect of fairness concerns into a risk-averse utility function.

2 Risk neutrality for fair wages

When the agent is risk-neutral over money, but has concerns for fairness, his utility function is given by:

$$u(w,q) \equiv w - \gamma \max\{w^*(q) - w, 0\}. \tag{8}$$

[3] This condition is sufficient but not necessary for a nonbinding IR.

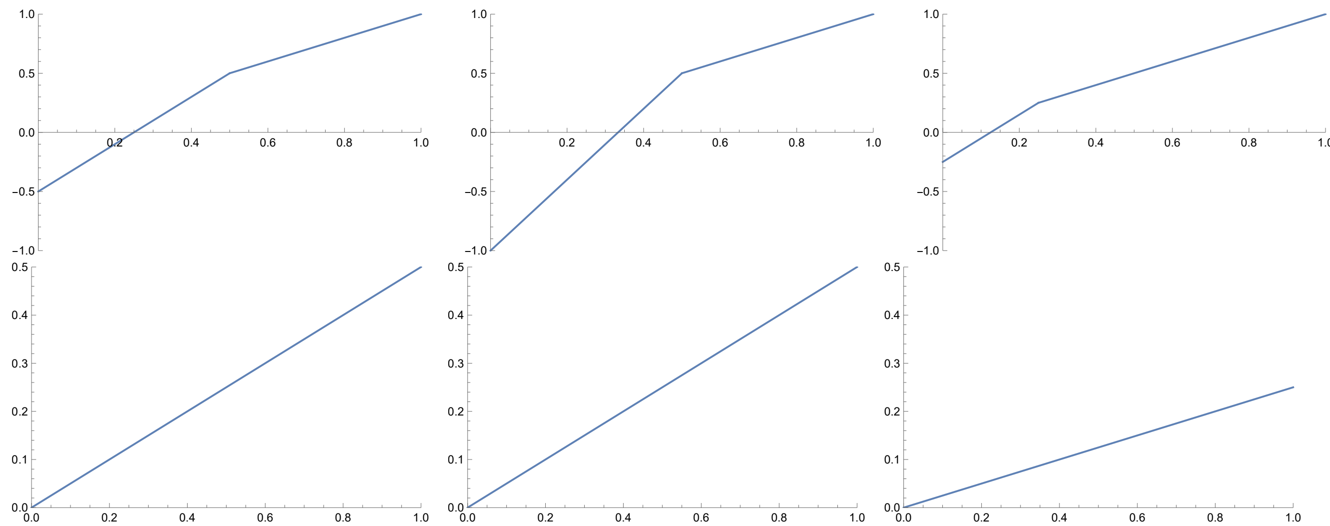

Figure 1: Top row: the function $u(w)$ defined in equation (8) as a function of w for $q = 1$, $\gamma = 1$ and $\rho = 0.5$ on the left; $\gamma = 2$ and $\rho = 0.5$ in the middle; $\gamma = 1$ and $\rho = 0.25$ on the right. Bottom row: the fair wage $w^*(q)$ defined in equation (9) as a function of q for $\gamma = 1$ and $\rho = 0.5$ on the left; $\gamma = 2$ and $\rho = 0.5$ in the middle; $\gamma = 1$ and $\rho = 0.25$ on the right.

The first term captures the agent's risk neutrality over money. The second term represents his concern for fairness, where $w^*(q)$ is the agent's perceived fair wage for output q, and $\gamma \geq 0$ parametrizes the intensity of his fairness concerns. Thus, if the agent's actual wage falls below his perceived fair wage, he suffers disutility loss. The fair wage is given by

$$w^*(q) \equiv \rho q, \tag{9}$$

where $\rho \geq 0$ is the agent's perceived fair share of output q.[4]

With $\gamma = 0$ (no fairness concerns) or $\rho = 0$ (any wage is perceived as fair), the utility function in equation (8) collapses to the standard risk-neutral utility function $u(w, q) = w$. Accordingly, unless otherwise specified, we assume $\gamma > 0$ and $\rho > 0$. Figure 1 illustrates the utility function for various values of γ and ρ. This piecewise linear utility function is the simplest and most transparent specification for fairness concerns, and allows us to conduct comparative statics with respect to the parameters γ and ρ.

As can be seen in Figure 1, the utility function exhibits loss aversion. The agent cares not only about the wage w per se, but also gains and losses relative to a reference point, with his sensitivity to losses exceeding his sensitivity to gains. De Meza and Webb (2007) and Dittmann, Maug, and Spalt (2010) also study optimal contracts in the presence of loss aversion. The main difference between fairness concerns and standard loss aversion is that, with the former, the reference point is the agent's perceived fair share of output and thus depends on output. Under standard loss

[4]One determinant of the perceived fair share is how much the agent can affect output, i.e. how sensitive output to effort. The survey of Edmans, Gosling, and Jenter (2022) finds that "how much the CEO can affect firm performance is the main determinant of pay variability" and the free-text fields and interviews suggest that fairness is a primary reason: "if the CEO has a greater effect on performance, it is fair to reward her more for good performance."

aversion, the reference point is independent of output. In de Meza and Webb (2007), it is the median of the wage distribution; in Dittmann, Maug, and Spalt (2010) it is last year's salary (they consider an alternative reference point that also includes the market value of the shares and options the agent inherited from the previous year.)

To simplify the analysis, in this section we assume:

$$-\gamma\rho\int_0^{q_0^{e^*}} q\frac{\partial}{\partial e}\phi(q|e^*)dq + \rho\int_{q_0^{e^*}}^{\overline{q}} q\frac{\partial}{\partial e}\phi(q|e^*)dq \geq C'(e^*) \tag{10}$$

$$-\gamma\rho\int_0^{\overline{q}} q\phi(q|0)dq < \overline{U} + C(0) \tag{11}$$

$$\rho\int_0^{\overline{q}} q\phi(q|e^*)dq \geq \overline{U} + C(e^*), \text{ where } e^* \text{ satisfies equation (2) with } w(q) = w^*(q)\ \forall q. \tag{12}$$

The assumptions in equations (10)-(12) are not crucial for our results; there would be other cases to consider without them. Equation (10) ensures that an incentive-compatible and individually rational contract that elicits effort e^* exists even if the firm never pays more than the fair wage. Define $q_m^{\min}$ implicitly as the highest value that satisfies the following equation:

$$\int_0^{q_m^{\min}} u(0,q)\frac{\partial}{\partial e}\phi(q|e^*)dq + \int_{q_m^{\min}}^{\overline{q}} u(w^*(q),q)\frac{\partial}{\partial e}\phi(q|e^*)dq \equiv C'(e^*). \tag{13}$$

If an incentive-compatible contract exists such that the agent is never paid above the fair wage for any output, $q_m^{\min}$ is the threshold such that the payment is zero below $q_m^{\min}$, and the fair wage above $q_m^{\min}$. Note that, with MLRP, the definition of $q_0^{e^*}$ and equation (13) imply that $q_m^{\min} \geq q_0^{e^*}$. Equation (11) implies that, even if the marginal cost of effort were zero, an agent who is paid zero for any output would be below his reservation utility and thus reject the contract. Equation (12) implies that a contract that always pays the fair wage and induces effort e^* satisfies the participation constraint, i.e. the agent accepts the contract.

Proposition 1 gives the optimal contract to implement a zero effort level, i.e. where the IC is slack and only goal of the contract is to ensure the agent's participation.

Proposition 1 *The contract that solves the optimization program in equations (1)-(4) with* $e^T = 0$ *is given by:*

$$w(q) = \begin{cases} 0 & \text{for } q < q_m \\ w^*(q) & \text{for } q \in [q_m, q_M] \end{cases} \tag{14}$$

where the thresholds q_m *and* q_M *are such that* $0 \leq q_m \leq q_M \leq \overline{q}$*. Moreover, when the agent's reservation utility is sufficiently large, this contract induces a strictly positive level of effort,* $e_0^* > 0$.

Proposition 2 gives the optimal contract when the IC binds, i.e. principal wants to induce a level of effort at least as high as e_0^*. It also characterizes the thresholds in the case in which the IR (3) is nonbinding.

Proposition 2 *For the program in equations (1)-(4) the optimal contract that induces a given effort* $e^T \geq e_0^*$ *is given by:*

$$w(q) = \begin{cases} 0 & \text{for } q < q_m \\ w^*(q) & \text{for } q \in [q_m, q_M] \\ q & \text{for } q > q_M \end{cases} \tag{15}$$

where the thresholds q_m *and* q_M *are such that* $0 \leq q_m \leq q_M \leq \overline{q}$*.*

(i) If $(1+\gamma)LR\left(q_m^{min}|e^*\right) > LR\left(\overline{q}|e^*\right)$ *and the participation constraint holds at* $q_m = q_m^{\min}$ *and* $q_M = \overline{q}$*, then* $q_m = q_m^{\min}$ *and* $q_M = \overline{q}$*, where* $q_m^{\min}$ *is implicitly defined in equation (13).*

(ii) Otherwise, then generically $q_m = q_m^\gamma$ *and* $q_M = q_M^\gamma$ *such that the contract is incentive compatible and individually rational and:*

$$(1+\gamma)LR\left(q_m^\gamma|e^*\right) + \frac{\eta_{IR}}{\eta_{IC}}\gamma = LR\left(q_M^\gamma|e^*\right), \tag{16}$$

where η_{IR} *and* η_{IC} *are the Lagrange multipliers associated with the IR and the IC, respectively (with* $\eta_{IR} = 0$ *when the IR is nonbinding).*

The optimal contract is discontinuous, and involves up to three regions. For $q < q_m$, the agent is paid zero. At $q = q_m$, the wage jumps to the fair wage $w^*(q)$. As output continues to rise, the fair wage $w^*(q) = \rho q$ rises and the agent continues to be paid his fair wage. Once output hits $q = q_M$, the wage jumps to the entire output.

The proof of Proposition 2 has two steps. First, it shows that an optimal contract takes the form given by equation (15). Indeed, for any subinterval of $[0, \overline{q}]$ with a payment in $(0, w^*(q))$, the contract can be improved by paying the agent either zero of the fair wage instead. Likewise, for any subinterval of $[0, \overline{q}]$ with a payment in $(w^*(q), q)$, the contract can be improved by paying the agent either the fair wage or the whole output instead. The contract then takes the form of equation (15) due to MLRP. Intuitively, higher payments are concentrated on outputs that are better indicators of the agent's effort.

The second step establishes the optimal values of q_m and q_M. In case (i) of Proposition 2, for any incentive-compatible q_m and q_M, the cost of the contract diminishes when q_M increases and q_m decreases correspondingly to retain incentive compatibility. Indeed, decreasing q_m allows to reduce the penalty for payments below the fair wage for outputs above $q_m^{\min}$, which have a positive likelihood ratio. When $(1+\gamma)LR\left(q_m^{\min}|e^*\right) > LR\left(\overline{q}|e^*\right)$, the penalty γ for payments below the fair wage is sufficiently strong for this effect to outweigh the standard desire to concentrate incentives on very high outputs (Innes (1990)). Thus, the optimal contract involves increasing q_M to the highest possible level of $\overline{q}$. Since this means that the agent is never paid above the fair wage for any output, incentive compatibility is achieved by setting $q_m = q_m^{\min}$ as in equation (13). In case (ii) of Proposition 2, when q_M is high and the corresponding incentive-compatible q_m is low (so that $(1+\gamma)LR\left(q_m|e^*\right) < LR\left(q_M|e^*\right)$), the cost of the contract diminishes when q_M decreases and q_m increases; on the other hand, when q_M is low and the corresponding incentive-compatible q_m is high (so that $(1+\gamma)LR\left(q_m|e^*\right) > LR\left(q_M|e^*\right)$), the cost of the contract diminishes when q_M

increases and q_m decreases. Thus, q_m and q_M are at intermediate rather than extreme levels, and are defined in equation (16).

We now discuss the effect of the intensity of fairness concerns, as measured by γ, on the contract. Without fairness concerns ($\gamma = 0$), we have $u(w, q) = w$ as in the pure moral hazard setting of Innes (1990). The principal generates effort incentives by rewarding high outputs; due to MLRP, it is efficient to concentrate rewards on very high outputs only. As a result, the optimal contract is "live-or-die" – with $\gamma = 0$, we have $q_m = q_M$; there is a single threshold and the agent is paid the minimum possible (zero) below the threshold and the maximum possible (the entire output q) above it. With fairness concerns ($\gamma > 0$), such a contract is suboptimal for two reasons. First, it does not satisfy the participation constraint efficiently, which is a concern if the participation constraint is binding. The agent is receiving an unfair wage (zero) for output levels below the threshold, which causes him disutility and may lead to him refusing the contract. Second, such a contract does not satisfy the incentive constraint efficiently. The agent is receiving an unfair wage for some output levels below the threshold, even though these output levels are associated with positive likelihood ratios. Thus, even though these output levels indicate that the agent has worked, he suffers significant disutility for achieving them, reducing his incentives to work. 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 are nevertheless associated with positive likelihood ratios) from 0 to $w^*(q)$, and simultaneously to reduce the rewards for moderately high outputs from q to $w^*(q)$.

Cases (ii) and (iii) of Proposition 2 establish that, when γ is positive but sufficiently low, the optimal contract has three regions. For outputs below q_m, the agent is paid zero; for outputs above q_M, the agent is paid the entire output. These regions are similar to Innes (1990). However, due to fairness concerns, there is a third region – for intermediate outputs, the agent is paid a fair wage. Thus, fairness concerns cause the agent to depart from the live-or-die contract and offer a fair wage for certain output levels. Part (i) shows that, when γ is sufficiently high, the highest region disappears, so the agent is never paid the entire output. The optimal contract thus only has two regions – he is paid zero for low outputs and the fair wage for high outputs. Intuitively, since fairness concerns are so strong, his utility rises much more from increasing his wage from 0 to $w^*(q)$ than from increasing it from $w^*(q)$ to q (per dollar increase). Thus, it is efficient for the principal to lower the rewards for very high outputs from q to $w^*(q)$, and use this to finance an increase in rewards for moderately low outputs with positive ratios from 0 to $w^*(q)$.

While the above explains the optimal contract by starting from a model of moral hazard and adding in fairness concerns, another way to view the intuition is to start with a pure fairness model and then add in moral hazard. One may think that fairness concerns would lead to the agent always being paid the fair wage $w^*(q)$, and this would indeed be the case if there were no moral hazard and so the only goal of the contract were to ensure the agent's participation. Since the agent suffers disutility from being paid an unfair wage, the principal provides the agent his reservation utility at minimum cost by always paying him is fair wage. However, such a contract does not provide effort incentives efficiently. In particular, 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. Thus, fairness concerns do not lead to fair wages for all outputs; in contrast, they can justify unfair wages for some outputs because avoiding unfairness is a motivator. In addition, if output is sufficiently high, the agent is paid the entire output even though this is more than required to meet his fairness constraint. This is because, for incentive provision, it is efficient to concentrate rewards in the highest likelihood ratio states; with a monotone likelihood ratio, this involved paying the agent the maximum possible for high outputs.

In case (i), the contract represents performance shares, where the agent is given shares in the firm worth ρq but these shares are forfeited if the output is below a threshold q_m. In standard models where the likelihood ratio is a continuous function of output (as in our setting), such as Holmstrom (1979), the optimal contract is also a continuous function of output and so does not involve discontinuities. In our model, discontinuities are optimal because, for moderate outputs, it is efficient to pay the agent his fair wage, but if output is sufficiently low, the agent is punished with the most unfair possible wage of zero, and this threat incentivizes effort. In Innes (1990) without a monotonicity constraint, the optimal contract is discontinuous but takes a "bang-bang" form where the agent receiving either the lowest possible wage or the highest possible wage; we are unaware of cases in which such a contract is offered in reality. Our contract involves discontinuities, but the discontinuity is non-extreme and leads to interior solutions – at q_m, the wage jumps from 0 to a share of output (rather than the entire output), as is the case with performance shares.

Innes (1990) obtains a realistic contract by assuming a monotonicity constraint – that the principal's payoff cannot be decreasing in output. Under such a constraint, the agent receives levered equity and the principal receives debt, and the contract has no discontinuities. Such a contract is realistic for a financing setting, in which the agent is an entrepreneur who is raising financing from the principal, an investor. However, it is less realistic for a contracting setting, in which the agent is a manager who is hired to work for the principal, the firm (or, more precisely, the shareholders of a firm). Hired managers are almost never residual claimants, who are given the entirety of the firm's equity. In addition, CEO contracts commonly involve discontinuities (see the survey of Edmans, Gabaix, and Jenter (2017) – executives are frequently offered bonuses if performance crosses above a certain level, or given equity which they forfeit if performance falls below a certain level. The two justifications in Innes (1990) for a monotonicity constraint also may be less applicable to an executive compensation setting. One justification is that, if the principal's payoff were decreasing, she would "burn" output. This might be possible in a financing setting where there is a single investor, but is difficult in a large public firm where there are dispersed shareholders. While the board acts on behalf of shareholders, "burning" output would be in violation of fiduciary duty. A second is that, if the agent's payoff increased more than one-for-one with output, he would secretly borrow to increase output. This may be possible in a financing setting in which output is cash flow, but in an executive compensation setting, the most relevant measure of output is the stock price. An executive secretly taking a loan to boost

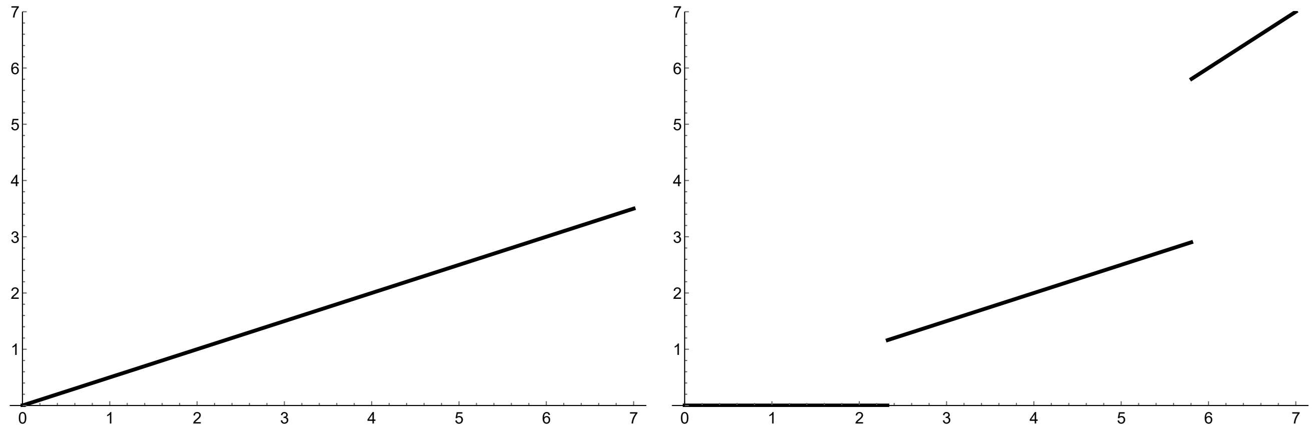

Figure 2: The contract $w(q)$ as a function of q for parameter values described in Example 1.

the stock price would likely be viewed as stock price manipulation. For all of these reasons, the monotonicity constraint is unlikely to relevant in an executive compensation setting, which explains why discontinuities are common in executive contracts. Our model generates an optimal contract that involves discontinuities, but also obtains a realistic contract without requiring a monotonicity constraint.

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

Corollary 1 *When the incentive constraint binds and $q_m \leq q_0^{e^T}$, the threshold q_m above which the manager is paid a fair wage $w^*(q)$ is decreasing in the intensity γ of fairness concerns.*

When incentive provision affects the design of the contract and the agent is paid a fair wage when output is "good news" about effort ($q_0^{e^T}$ is the level of output such that the likelihood ratio is zero), the threshold above which the manager is paid a fair wage $w^*(q)$ is decreasing in the intensity of fairness concerns, as measured by γ. The stronger fairness concerns are, the more important it is to pay fair wages, and so the principal does so over a larger range of outputs.

Example 1 illustates the agent's preferences and the optimal contract for a given parametrization.

Example 1 *In this numerical example, the agent's preferences are described by $\gamma = 1$, $\rho = \frac{1}{2}$, and $C(e) = c \times e^2$. Output is lognormally distributed with parameters $e^* = 1$ and $\sigma = 1$. The optimal contract is depicted on the left of Figure 2 for $c = 1$ and $\overline{U} = 2$, and on the right of Figure 2 for $c = \frac{3}{2}$ and $\overline{U} = \frac{3}{2}$.*

3 Risk aversion for fair wages

In this section, the utility function is defined as:

$$u(w, q) \equiv \min \{v(w), \nu(w, q)\} \tag{17}$$

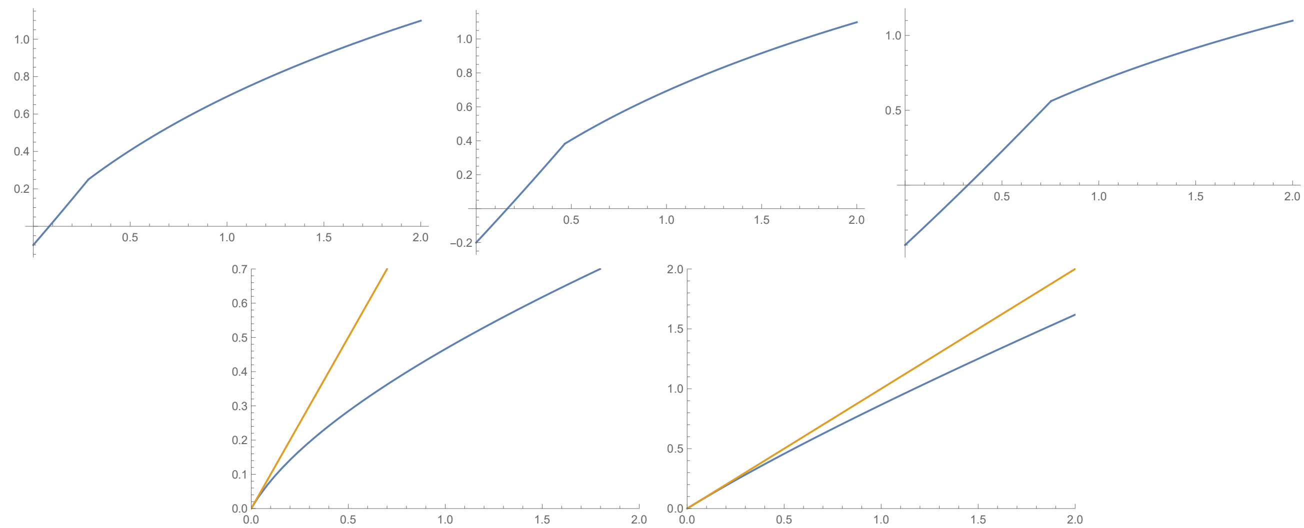

Figure 3: Top row: the function $u(w)$ defined in equation (17) as a function of w for $v(w) = \ln(w+1)$ and $\nu(w,q) = (w+1)^{1.2} - 1 - \frac{1}{5}q$ with $q = 0.5$ on the left, $q = 1$ in the middle, and $q = 2$ on the right. Bottom row: the blue line is the fair wage $w^*(q)$ defined by $v(w^*(q)) \equiv \nu(w^*(q), q)$ 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, and $v(w) = \sqrt{w+1} - 1$ and $\nu(w,q) = w - \frac{1}{2}q$ on the right. The orange line is principal LL.

where $v(w)$ is the utility over money alone, which is increasing and concave ($v' > 0$, $v'' \leq 0$), with $v(0) = 0$.[5] 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 have $\nu(0,q) \leq 0$, $\nu'_q(w,q) < 0$ (higher output raises the fair wage and thus lowers utility), $\nu'_w(w,q) > 0$, $\nu''_{ww}(w,q) \geq 0$, $\nu''_{wq} = 0$, and $\nu''_{qq} = 0$. For any given q, the two functions $v(w)$ and $\nu(w,q)$ intersect on $(0,\infty)$ at most once.[6] Let this point, if it exists, be denoted by $w^*(q)$, which is such that $v(w^*(q)) \equiv \nu(w^*(q), q)$. At this point, $v'(w) < \nu'(w,q)$, so that, for a given q, there is a kink in the utility function $u(w,q)$ as a function of w at $w = w^*(q)$. Thus, $w^*(q)$ captures the agent's perceived fair wage, as in Section 2. This utility function (17) exhibits not only loss aversion, but also concavity above the reference point $w^*(q)$ and convexity below it, as in prospect theory.[7]

We also assume that for any q,

$$\lim_{w \searrow 0} \nu'(w,q) > \lim_{w \searrow 0} v'(w) \tag{18}$$

so that the utility function is always steeper below the fair wage than above the fair wage.

[5] This specification for the function $v(w)$ includes CRRA utility with relative risk aversion less than 1, and a "normalized" version of log utility $v(w) = \ln(w+1)$.

[6] 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.

[7] However, the utility function (17) does not exhibit probability weighting as in prospect theory.

To guarantee existence of an optimal contract, we assume that:

$$\int_0^{q_0^{e^*}} u(w^*(q), q) \frac{\partial}{\partial e} \phi(q|e^*) dq + \int_{q_0^{e^*}}^{\overline{q}} u(q, q) \frac{\partial}{\partial e} \phi(q|e^*) dq > C'(e^T) \tag{19}$$

$$\int_0^{q_0^{e^*}} u(w^*(q), q) \phi(q|e^*) dq + \int_{q_0^{e^*}}^{\overline{q}} u(q, q) \phi(q|e^*) dq - C(e^*) > \overline{U} \tag{20}$$

Equations (19) and (20) imply that, given the endogenous effort choice, an agent who receives the fair wage for outputs with a negative likelihood ratio and the whole output for a positive likelihood ratio with choose a level of effort of at least e^T and will be above his reservation utility. We also assume that an agent who is paid his fair wage for any output is at his reservation utility:

$$\int_0^{\overline{q}} u(w^*(q), q) \phi(q|e^F) dq - C(e^F) \geq \overline{U}, \tag{21}$$

where e^F be the effort level induced when the agent is paid his fair wage for any output:

$$\int_0^{\overline{q}} u(w^*(q), q) \frac{\partial}{\partial e} \phi(q|e^F) dq = C'(e^F). \tag{22}$$

The following assumption ensures that a manager who is always paid zero is below his reservation utility:

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

To better understand the effects of incentive provision on the optimal contract, we start by considering the relaxed program when the target level of effort e^T is zero, i.e., the principal does not consider incentive provision when designing the contract. Thus, the principal chooses the contract to minimize the cost of ensuring the agent's participation subject to double-sided limited liability.

Proposition 3 *The contract that solves the optimization program in equations (1)-(4) with* $e^T = 0$ *is given by:*

$$w(q) = \begin{cases} 0 & \text{for } q < q_m \\ w^*(q) & \text{for } q \in [q_m, q_M] \end{cases}. \tag{23}$$

Proposition 3 highlights an interesting implication of fairness concerns. Even if there were no moral hazard, the optimal contract need not involve a fixed wage. In turn, a wage that is both increasing in output and also fair provides incentives to exert positive effort, and so the principal obtains effort "for free". Note that it is insufficient for the wage to be merely increasing in output for it to provide effort incentives – it may fail to do so if it is also unfair. Indeed, since $\nu(w, q)$ is decreasing in q, a payment schedule which is only slightly increasing in output and is below the fair wage may fail to elicit effort. Intuitively, even though higher output increases the wage, for an increasing wage schedule, it also increases the agent's reference point. Thus, if the reference point increases by more than the actual wage, the agent's utility does not increase, and so he is not rewarded for increasing output.

The result in Proposition 4 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. Without an incentive constraint, the optimal contract involve a fixed wage for optimal risk-sharing; inducing effort requires an output-contingent wage which leads to deviation from efficient risk-sharing and is thus costly. As a result, any effort level in $[0, \bar{e}]$ can in principle be optimal, depending on model parameters. As we will show in Proposition 5, this is not true in a model with fairness concerns.

Proposition 4 *When $e^T = 0$ and equation (21) holds as an equality, the contract that solves the optimization problem in equations (1)-(4) is $w^*(q)$. Moreover, this contract induces a strictly positive level of effort.*

Proposition 4 states that, when the agent is at his reservation level of utility when he receives a fair wage and the principal does not consider effort provision when designing the contract, the cheapest contract to give the agent an expected utility of $\overline{U}$ involves paying the agent his fair wage $w^*(q)$ for any output.

Proposition 5 *When $e^T = 0$ and equation (21) holds as an equality, any effort level lower than e^F is suboptimal and will not be induced by the principal.*

Proposition 5 states that a principal that faces the same problem as in Proposition 4 will never induce an effort level less than e^F. Intuitively, providing low effort incentives either requires paying unfair wages for high outputs (which reduces expected utility) or paying in excess of fair wages for low outputs (which is unnecessarily costly). Critics of high pay-performance sensitivity argue that it is not needed to incentivize effort, since boards should monitor effort or CEOs should be intrinsically motivated. However, performance-sensitive contracts may be offered not to provide incentives, but to ensure the CEO is fairly-paid. A by-product of fair pay is that it incentivizes effort, even if such incentives are unnecessary due to board monitoring or intrinsic motivation. Without fairness concerns, it is costly to incentivize high effort levels; with fairness concerns, it is costly to incentivize low effort levels as doing so requires offering unfair pay.

Proposition 6 considers the general case in which both the IC and IR may be binding or nonbinding. The variables λ_{IR} and λ_{IC} denote the Lagrange multipliers associated with the IR and IC, respectively.

Proposition 6 *For the program (1)-(4), the optimal contract is such that, for $q_m \in [0, \overline{q}]$:*

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

The optimal contract is given by four regions. As in the case of risk-neutrality, there are three regions in which the agent is paid zero, the fair wage, and the entire output. However, there is a fourth region, given by $q \in (q_M, q_N)$. For these output levels, output is sufficiently high that the principal wishes to reward the agent by more than his fair wage ($v'^{-1}\left(1/\left(\lambda_{IR} + \lambda_{IC} LR(q|e^*)\right)\right) > w^*(q)$). However, it is inefficient to give him the entire output, since the agent exhibits diminishing marginal utility and so does not value this additional reward highly. Thus, unlike in the risk-neutral case, the optimal contract is continuous at q_M – the principal pays the agent more than his fair wage, but not the entire output. 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.[8] For $q > q_N$, the likelihood ratio is so high that the principal rewards the agent with the entire output.

A special case of Proposition 6 is as follows. With "normalized" log utility ($v(w) = \ln(w + 1)$)[9] and a normally distributed output with mean e^* and variance σ^2, $LR(q|e^*) = \frac{q-e^*}{\sigma^2}$, and $v'^{-1}\left(1/\left(\lambda_{IR} + \lambda_{IC} LR(q|e^*)\right)\right)$ is linear in q. Thus, the contract is piecewise linear with a slope of 0 in the first region, ρ in the second, a number between ρ and 1 in the third, and 1 in the fourth.

When the participation constraint is nonbinding, the payment in the third region (where the agent is paid more than the fair wage but less than the entire output) is a linear, as opposed to affine, function of the likelihood ratio. This implies that the payment is positive if and only if the likelihood ratio is positive. When the participation constraint is binding ($\lambda_{IR} \geq 0$), tthe payment is instead an affine function of the likelihood ratio and so it can be positive even if the likelihood ratio is negative (i.e. $q < q_0^{e^*}$). Intuitively, paying the agent zero for outputs below $q_0^{e^*}$ increases incentives, but significantly reduces the agent's utility due to the perceived unfairness, and may lead to his participation constraint being violated.

Proposition 6 also shows that, perhaps surprisingly, pay is increasing in output even when the IC is non-binding ($\lambda_{IC} = 0$) and so an increasing wage is not needed to incentivize effort. The intuition has two parts. First, the fair wage is increasing in output, and paying the agent his fair wage for a large subset of outputs helps satisfy the participation constraint. However, the participation constraint can be satisfied even if the agent is not paid a fair wage for all outputs. The question then becomes how to allocate payments at the fair wage and below the fair wage across output realizations. With $\nu'' > 0$, the agent's utility is non-concave for payments below the fair wage. Thus, it is efficient to pay the agent either zero or the fair wage rather than an intermediate level. Since the fair wage is increasing in output, the disutility from zero wages is also increasing in output, and so it is optimal to pay zero wages for low output levels.

Proposition 6 shows how whether the participation constraint binds affects the optimal contract.

[8] 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 means that protecting the agent against downside risk is relatively unimportant, but providing strong incentives where the agent's marginal utility is high (i.e. for low outputs) is especially important. It will typically not be satisfied for CEOs who have low relative risk aversion due to their wealth.

[9] In this case, $v'(w) = \frac{1}{w+1}$ $v'^{-1}(w) = \frac{1}{w} - 1$, so that $v'^{-1}\left(1/\lambda LR(q|e^*)\right) = LR(q|e^*) - 1$.

When the participation constraint does not bind ($\lambda_{IR} = 0$), we have $q_m > 0$, so that the contract has a discontinuity between zero payments and positive payments. It is optimal to pay the agent the most unfair feasible wage for low outputs, to incentivize him to exert effort and avoid low outputs. However, doing so risks the agent suffering significant disutility and may fail to ensure the agent's participation. Thus, when the participation constraint binds, the contract may not have a discontinuity (see Proposition 4). Overall, the participation constraint binding is a necessary but insufficient condition for pay to be a continuous function of output. As a result, the increasing use of performance shares, which do contain discontinuities, 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.

References

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

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

[3] Charness, Gary and Matthew Rabin (2002): "Understanding Social Preferences With Simple Tests." *Quarterly Journal of Economics* 117, 817-869

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

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

[6] 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.

[7] Edmans, Alex, Tom Gosling, and Dirk Jenter (2022): "CEO Compensation: Evidence From the Field." Working Paper, London Business School.

[8] Fehr, Ernst, Georg Kirchsteiger, and Arno Riedl (1993): "Does Fairness Prevent Market Clearing? An Experimental Investigation." *Quarterly Journal of Economics* 108, 437–459.

[9] Fehr, Ernst, Alexander Klein, and Klaus M. Schmidt (2007): "Fairness and Contract Design." *Econometrica* 75, 121–154.

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

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

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

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

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

[15] Rabin, Matthew (1993): "Incorporating Fairness into Game Theory and Economics." *American Economic Review* 83, 1281–1302.

[16] 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:

$$\int_0^{\overline{q}} u\left(w(q), q\right) \frac{\partial^2 \phi(q|e)}{\partial e^2} dq < C''(e) \quad \forall e \in (0, \bar{e}). \tag{24}$$

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^+ and K_e^- defined in equations (5) and (6), the expression on the left-hand side ("LHS") of equation (24) can then be rewritten as:

$$\int_0^{\overline{q}} u\left(w(q), q\right) \min\left\{\frac{\partial^2 \phi(q|e)}{\partial e^2}, 0\right\} dq + \int_0^{\overline{q}} u\left(w(q), q\right) \max\left\{\frac{\partial^2 \phi(q|e)}{\partial e^2}, 0\right\} dq. \tag{25}$$

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 $\frac{\partial^2 \phi(q|e)}{\partial e^2} \leq 0$ we have $u\left(w(q), q\right) \frac{\partial^2 \phi(q|e)}{\partial e^2} \leq u(0, q) \frac{\partial^2 \phi(q|e)}{\partial e^2}$; and for any q such that $\frac{\partial^2 \phi(q|e)}{\partial e^2} \geq 0$ we have $u\left(w(q), q\right) \frac{\partial^2 \phi(q|e)}{\partial e^2} \leq u(q, q) \frac{\partial^2 \phi(q|e)}{\partial e^2}$. Integrating over q, this implies that expression (25) is less than:

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

which completes the proof.

Proof of Proposition 1:

We describe the optimal contract when $e^T = 0$, i.e. the IC does not bind. 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) and (4) is simply $w(q) = 0$ for any q, and $u(0, q) = -\gamma \max\{\rho q, 0\} = -\gamma \rho q$ for any $q \in [0, \overline{q}]$, so that, using equation (11) with

$C' > 0$ and $e^* \geq 0$:

$$\int_0^{\overline{q}} u(0,q)\phi(q|e^*)dq = -\gamma\rho\int_0^{\overline{q}} q\phi(q|e^*)dq < 0 \quad \Rightarrow \quad -\gamma\rho\int_0^{\overline{q}} q\phi(q|e^*)dq - C(e^*) < \overline{U}$$

i.e. the IR is not satisfied, a contradiction.

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

$$\min_{w(q)} \int_0^{\overline{q}} w(q)\phi(q|e^*)dq \tag{26}$$

$$\text{s.t.} \quad \int_0^{\overline{q}} u(w(q),q)\phi(q|e^*)dq - C(e^*) = \overline{U} \tag{27}$$

$$0 \leq w(q) \leq q \tag{28}$$

Lemma 2 *Let the utility function be as in equation (8) and suppose that the IC is nonbinding. On any non-empty subinterval of* $[0,\overline{q}]$*, the optimal contract is such that* $w(q) \leq w^*(q)$*.*

Proof. This proof is by contradiction.

A contract such that $w(q) \geq w^*(q)$ for all $q \in [0,\overline{q}]$ would not solve the optimization program in equations (26)-(28). Indeed, due to equation (12) the agent would be strictly above his reservation utility so that the payment $w(q)$ could be reduced on some subinterval of $[0,\overline{q}]$, which would decrease the cost of the contract in equation (26) without violating the IR in equation (27) or the limited liability constraints in equation (28), a contradiction. Thus, any optimal contract is such that $w(q) < w^*(q)$ for some q. For a given contract, denote the subset of $[0,\overline{q}]$ such that $w(q) < w^*(q)$ by Q^-.

We now show that a contract with $w(q) > w^*(q)$ for any $q \in [0,\overline{q}]$ is suboptimal. Suppose that we have $w(q) \in (w^*(q),q]$ on a non-empty subinterval of $[0,\overline{q}]$, which we denote by Q^+. Consider the following perturbation for $q \in Q^-$, increase $w(q)$ by $\epsilon/\phi(q|e^*)$, and for q'^+ decrease $w(q')$ by $\epsilon/\phi(q'^*)$, where ϵ is positive and arbitrarily small. By construction, this perturbation is cost-neutral for a given effort, i.e. it does not change the principal's objective function. Now consider the effect on the LHS of the IR in equation (27). Since $w(q) \in [0,w^*(q))$ and $w(q'^*(q'),q']$, the change in the LHS of the IR is: $\epsilon(1+\gamma) - \epsilon = \epsilon\gamma$, which is strictly positive since $\gamma > 0$. Since the LHS of the IR increases and the IR is binding, standard arguments show that it is then possible to construct a contract that leaves the LHS of the 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 $w(q) > w^*(q)$ for some q.

■

By using Lemmas 5 and 6, which hold for $v'' \leq 0$ and $\nu'' \geq 0$ including the special case of the utility function in equation (8), and Lemma 2 above, which holds for a utility function as in equation (8), we get that the optimal contract when the IC is nonbinding takes the form:

$$w(q) = \begin{cases} 0 & \text{if } q < q_m \\ w^*(q) & \text{if } q \geq q_m \end{cases} \tag{29}$$

For a contract as in equation (29) and an equilibrium effort e^* given by equation (2) with $e^T = 0$, define:

$$A(q_m, e) \equiv -\gamma\rho \int_0^{q_m} q \frac{\partial}{\partial e}\phi(q|e)dq + \rho \int_{q_m}^{\overline{q}} q \frac{\partial}{\partial e}\phi(q|e)dq \tag{30}$$

Given the FOA, for a given q_m the equilibrium effort is zero if and only if $A(q_m, 0) \leq 0$. Since the IR is binding and the FOA applies for any interior solution to the effort choice problem, the optimal value of q_m and the effort induced, e^*, are implicitly defined by:

$$-\gamma\rho \int_0^{q_m} q\phi(q|e^*)dq + \rho \int_{q_m}^{\overline{q}} q\phi(q|e^*)dq - C(e^*) = \overline{U} \tag{31}$$

$$e^* = \begin{cases} 0 & \text{if } A(q_m, 0) \leq 0 \\ C'^{-1}\left(A(q_m, e^*)\right) & \text{if } A(q_m, 0) > 0 \end{cases} \tag{32}$$

As $q_m \to \overline{q}$, since $\int_0^{\overline{q}} \frac{\partial}{\partial e}\phi(q|e)dq = 0$ for any e by definition of a density, $A(q_m, e)$ is negative for any e, i.e. the equilibrium effort is zero by equation (32). For $q_m = 0$, $A(q_m, e)$ has the same sign as $\int_0^{\overline{q}} q\frac{\partial}{\partial e}\phi(q|e)dq$, which is strictly positive for any e because of $\int_0^{\overline{q}} \frac{\partial}{\partial e}\phi(q|e)dq = 0$ for any e and MLRP. Therefore, $A(0, e) > 0$ for any e, including $e = 0$, i.e. $e^* > 0$ for $q_m = 0$. Moreover, denoting by e^* the equilibrium effort given the contract with threshold q_m, the derivative of the right-hand side ("RHS") of equation (30) with respect to q_m holding effort constant is:

$$-\gamma\rho q_m \frac{\partial}{\partial e}\phi(q_m|e^*) - \rho q_m \frac{\partial}{\partial e}\phi(q_m|e^*),$$

which is strictly positive if and only if $q_m < q_0^{e^*} \Leftrightarrow \frac{\partial}{\partial e}\phi(q_m|e^*) < 0$. In sum, $e^* > 0$ if $q_m \leq \hat{q}_m \in [q_0^{e^*}, \overline{q})$. Since $C'(0) = 0$ and $C'' > 0$, this implies that $e^* > 0$ if $q_m \leq \hat{q}_m$.

Consider the effect of a change in q_m on the LHS of the IR in equation (31) when $e^* > 0$ and is therefore given by the first-order condition ("FOC") to the agent's effort choice problem given the FOA. The derivative of the LHS of equation (31) with respect to q_m holding effort constant is:

$$-\gamma\rho q_m \phi(q_m|e) - \rho q_m \phi(q_m|e),$$

which is strictly negative for any e. Moreover, the LHS of the IR in equation (31) is the agent's objective function, and the agent chooses effort e to maximize this objective function. From the envelope theorem, we know that the total effect of a change in a parameter (here q_m) on the objective function is equal to its effect holding effort constant. In sum, the LHS of equation (31) is continuously decreasing in q_m, it is below the RHS for $q_m = \overline{q}$ according to equation (11), and it is above the RHS for $q_m = 0$ according to equation (12). This implies that there exists q_m that satisfies equation (31), and this level of q_m is strictly decreasing in $\overline{U}$ according to equation (31). In sum, when $\overline{U}$ is sufficiently high, we have $q_m < \hat{q}_m$, and the equilibrium level of effort is strictly positive according to the previous paragraph.

Proof of Proposition 2:

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

$$\int_0^{\overline{q}} u(w(q), q)\frac{\partial}{\partial e}\phi(q|e^*)dq = C'^*), \tag{33}$$

with $e^* = e^T$. We now describe the optimal contract when the IC in equation (2) binds.

The first step of the proof establishes that a contract as described in equation (15) is optimal. To this end, we rely on Lemmas 5 and 6, which hold for $v'' \leq 0$ and $\nu'' \geq 0$ including the special case of the utility function in equation (8), and Lemma 3, which holds for a utility function as in equation (8).

Lemma 3 *Let the utility function be as in equation (8) and suppose that the IC is binding. On any non-empty subinterval of* $[0, \overline{q}]$*, the optimal contract is such that* $w(q) \notin (w^*(q), q)$*.*

Proof. This proof is by contradiction. Suppose that, on a non-empty subinterval of $[0, \overline{q}]$, 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 $\epsilon/\phi(q'^*)$, where ϵ is positive and arbitrarily small. By construction, this perturbation is cost-neutral for a given effort, i.e. it does not change the principal's objective function. Now consider the effect on the LHS of the IC in equation (33). With $w(q) \in (w^*(q), q)$ and $w(q'^*(q'), q')$, the change in the LHS of the IC is:

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

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 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 $w(q) \in (w^*(q), q)$ on any non-empty subinterval of $[0, \overline{q}]$.

■

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

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

$$\min_{q_m, q_M} \int_0^{\overline{q}} w(q)\phi(q|e^*)dq \tag{34}$$

$$\text{s.t.} \quad \int_0^{\overline{q}} u(w(q), q)\frac{\partial}{\partial e}\phi(q|e^*)dq = C'^*) \tag{35}$$

$$\int_0^{\overline{q}} u(w(q), q)\phi(q|e^*)dq - C(e^*) \geq \overline{U} \tag{36}$$

$$w(q) = \begin{cases} 0 & \text{for } q < q_m \\ w^*(q) & \text{for } q \in [q_m, q_M] \\ q & \text{for } q > q_M \end{cases} \tag{37}$$

With the utility function defined in equation (8), this can be rewritten as, for $q_m \in [0, \overline{q}]$ and $q_M \in [q_m, \overline{q}]$:

$$\min_{q_m, q_M} \int_{q_m}^{q_M} \rho q \phi(q|e^*)dq + \int_{q_M}^{\overline{q}} q\phi(q|e^*)dq \tag{38}$$

$$\text{s.t.} \int_0^{q_m} (-\gamma\rho q)\frac{\partial}{\partial e}\phi(q|e^*)dq + \int_{q_m}^{q_M} \rho q\frac{\partial}{\partial e}\phi(q|e^*)dq + \int_{q_M}^{\overline{q}} q\frac{\partial}{\partial e}\phi(q|e^*)dq = C'^*) \tag{39}$$

$$\int_0^{q_m} (-\gamma\rho q)\,\phi(q|e^*)dq + \int_{q_m}^{q_M} \rho q\phi(q|e^*)dq + \int_{q_M}^{\overline{q}} q\phi(q|e^*)dq - C(e^*) \geq \overline{U} \tag{40}$$

Denote by η_{IC} and η_{IR} the Lagrange multipliers associated with the constraints in equations (39) and (40), respectively. The FOC for an interior solution are:

$$\begin{aligned} -\rho q_m\phi(q_m|e^*) - \eta_{IC}\left(-\gamma\rho q_m\frac{\partial}{\partial e}\phi(q_m|e^*) - \rho q_m\frac{\partial}{\partial e}\phi(q_m|e^*)\right) \\ -\eta_{IR}\left(-\gamma\rho q_m\phi(q_m|e^*) - \rho q_m\phi(q_m|e^*)\right) = 0 \end{aligned} \tag{41}$$

$$\begin{aligned} \rho q_M\phi(q_M|e^*) - q_M\phi(q_M|e^*) - \eta_{IC}\left(\rho q_M\frac{\partial}{\partial e}\phi(q_M|e^*) - q_M\frac{\partial}{\partial e}\phi(q_M|e^*)\right) \\ -\eta_{IR}\left(\rho q_M\phi(q_M|e^*) - q_M\phi(q_M|e^*)\right) = 0 \end{aligned} \tag{42}$$

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

$$-1 + \eta_{IC}\frac{\frac{\partial}{\partial e}\phi(q_m|e^*)}{\phi(q_m|e^*)}(1+\gamma) + \eta_{IR}(1+\gamma) = 0 \tag{43}$$

$$-1 + \eta_{IC}\frac{\frac{\partial}{\partial e}\phi(q_M|e^*)}{\phi(q_M|e^*)} + \eta_{IR} = 0 \tag{44}$$

The optimal value of q_m is not described by a corner solution. Indeed, we cannot have $q_m = 0$ or $q_m = \overline{q}$, which would imply $q_M = \overline{q}$ (this violates IC since the LHS is then negative and the RHS positive; this also violates IR according to equation (11)). Thus, the optimal value of q_m is given by the first-order condition in equation (43), 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 $e^* > 0$ must be binding. Suppose that it is not. Then, the contract that solves the optimization problem in equations (1) and (4) is simply $w(q) = 0$ for any q, and $u(0,q) = -\gamma \max\{\rho q, 0\} = -\gamma\rho q$ for any $q \in [0, \overline{q}]$, so that:

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

i.e. the IC is not satisfied, a contradiction. Thus, the IR is nonbinding, we have $\eta_{IR} = 0$ and the IC must be binding.

If the optimal values of q_m and q_M are interior solutions, equations (43) and (44) immediately give:

$$\frac{\frac{\partial}{\partial e}\phi(q_m|e^*)}{\phi(q_m|e^*)}(1+\gamma) = \frac{\frac{\partial}{\partial e}\phi(q_M|e^*)}{\phi(q_M|e^*)} \tag{45}$$

With a nonbinding IR, we establish that $q_m > q_0^{e^*}$, where $e^* = e^T$. For any $q \leq q_0^{e^*}$, consider any given initial compensation contract such that $w(q) > 0$ and the following perturbation: reduce $w(q)$ by an arbitrarily small amount ϵ. 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 (35) is continuously differentiable in q_m and q_M, and the product of continuous functions is continuous. Combining this with the results from the preceding paragraph, this shows that $\mathcal{Q}^{IC}$ is non-empty, and we have $\frac{dq_M^{IC}}{dq_m^{IC}} < 0$ with a nonbinding IR.

Totally differentiating the LHS of the IC with respect to q_m^{IC} and adjusting q_M^{IC} so that the LHS of the IC remains unchanged gives:

$$\begin{aligned}
\frac{d}{dq_m^{IC}}\int_0^{\overline{q}} u(w(q),q)\frac{\partial}{\partial e}\phi(q|e^*)dq &= \left(u(0,q_m^{IC}) - u(w^*(q_m^{IC}),q_m^{IC})\right)\frac{\partial}{\partial e}\phi(q_m^{IC}|e^*) \\
&\quad + \left(\left(u(w^*(q_M^{IC}),q_M^{IC}) - u(q_M^{IC},q_M^{IC})\right)\frac{\partial}{\partial e}\phi(q_M^{IC}|e^*)\right)\frac{dq_M^{IC}}{dq_m^{IC}} \\
&= -(1+\gamma)w^*(q_m^{IC})\frac{\partial}{\partial e}\phi(q_m^{IC}|e^*) - (q_M^{IC} - w^*(q_M^{IC}))\frac{\partial}{\partial e}\phi(q_M^{IC}|e^*)\frac{dq_M^{IC}}{dq_m^{IC}} = 0 \\
&\Leftrightarrow \qquad \frac{dq_M^{IC}}{dq_m^{IC}} = -\frac{(1+\gamma)w^*(q_m^{IC})}{q_M^{IC} - w^*(q_M^{IC})}\frac{\frac{\partial}{\partial e}\phi(q_m^{IC}|e^*)}{\frac{\partial}{\partial e}\phi(q_M^{IC}|e^*)} \qquad (46)
\end{aligned}$$

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 (34) unchanged for the principal. By construction:

$$\begin{aligned}
\frac{d}{dq_m^c}\int_0^{\overline{q}} w(q)\phi(q|e^*)dq = -w^*(q_m^c)\phi(q_m^c|e^*) - (q_M^c - w^*(q_M^c))\phi(q_M^c|e^*)\frac{dq_M^c}{dq_m^c} = 0 \\
\Leftrightarrow \quad \frac{dq_M^c}{dq_m^c} = -\frac{w^*(q_m^c)}{q_M^c - w^*(q_M^c)}\frac{\phi(q_m^c|e^*)}{\phi(q_M^c|e^*)} \qquad (47)
\end{aligned}$$

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

$$\frac{\frac{\partial}{\partial e}\phi(q_m|e^*)}{\phi(q_m|e^*)} \leq \frac{\frac{\partial}{\partial e}\phi(q_M|e^*)}{\phi(q_M|e^*)} \quad \Leftrightarrow \quad \frac{\frac{\partial}{\partial e}\phi(q_m|e^*)}{\frac{\partial}{\partial e}\phi(q_M|e^*)} \leq \frac{\phi(q_m|e^*)}{\phi(q_M|e^*)}, \qquad (48)$$

with strict inequalities for $q_M > q_m$. Likewise, because of MLRP, for $q_M \geq q_0^{e^*} > q_m$, we have:

$$\frac{\frac{\partial}{\partial e}\phi(q_m|e^*)}{\phi(q_m|e^*)} < \frac{\frac{\partial}{\partial e}\phi(q_M|e^*)}{\phi(q_M|e^*)} \quad \Leftrightarrow \quad \frac{\frac{\partial}{\partial e}\phi(q_m|e^*)}{\frac{\partial}{\partial e}\phi(q_M|e^*)} > \frac{\phi(q_m|e^*)}{\phi(q_M|e^*)}. \qquad (49)$$

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 $(1+\gamma)\frac{\frac{\partial}{\partial e}\phi(q_m^{IC}|e^*)}{\phi(q_m^{IC}|e^*)} < \frac{\frac{\partial}{\partial e}\phi(q_M^{IC}|e^*)}{\phi(q_M^{IC}|e^*)}$ 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 (46) results in a lower cost to the principal because of equations (47) and (48).

2) For values of q_m^{IC} and q_M^{IC} such that $(1+\gamma)\frac{\frac{\partial}{\partial e}\phi(q_m^{IC}|e^*)}{\phi(q_m^{IC}|e^*)} > \frac{\frac{\partial}{\partial e}\phi(q_M^{IC}|e^*)}{\phi(q_M^{IC}|e^*)}$ 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 (46) results in a higher cost to the principal because of equations (47) and (48).

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 (10), 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 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 (13). From equations (10) and (13), we have $q_m^{\min} \geq q_0^{e^*}$.

First, if $(1+\gamma)\frac{\frac{\partial}{\partial e}\phi(q_m^{\min}|e^*)}{\phi(q_m^{\min}|e^*)} > \frac{\frac{\partial}{\partial e}\phi(\overline{q}|e^*)}{\phi(\overline{q}|e^*)}$, then due to MLRP and $\frac{dq_M^{IC}}{dq_m^{IC}} < 0$, for any element of $\mathcal{Q}^{IC}$, we have $(1+\gamma)\frac{\frac{\partial}{\partial e}\phi(q_m^{IC}|e^*)}{\phi(q_m^{IC}|e^*)} > \frac{\frac{\partial}{\partial e}\phi(q_M^{IC}|e^*)}{\phi(q_M^{IC}|e^*)}$, 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:

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

where $q_m^{\min}$ is defined in equation (13).

Second, if $(1+\gamma)\frac{\frac{\partial}{\partial e}\phi(q_m^{\min}|e^*)}{\phi(q_m^{\min}|e^*)} < \frac{\frac{\partial}{\partial e}\phi(\overline{q}|e^*)}{\phi(\overline{q}|e^*)}$, 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 are respectively q_m^{γ} and q_M^{γ}, which belong to the subset $\mathcal{Q}^{IC}$ and satisfy the following equation:

$$(1+\gamma)\frac{\frac{\partial}{\partial e}\phi(q_m^{\gamma}|e^*)}{\phi(q_m^{\gamma}|e^*)} = \frac{\frac{\partial}{\partial e}\phi(q_M^{\gamma}|e^*)}{\phi(q_M^{\gamma}|e^*)}. \tag{51}$$

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

$$-\gamma\rho\int_0^{q_m} q\frac{\partial}{\partial e}\phi(q|e^*)dq + \rho\int_{q_m}^{q_M} q\frac{\partial}{\partial e}\phi(q|e^*)dq + \int_{q_M}^{\overline{q}} q\frac{\partial}{\partial e}\phi(q|e^*)dq = C'^*) \tag{52}$$

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

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

1. First, if the optimal values of q_m and q_M are interior solutions, equations (43) and (44) immediately give:

$$\frac{\frac{\partial}{\partial e}\phi(q_m|e^*)}{\phi(q_m|e^*)}(1+\gamma) + \frac{\eta_{IR}}{\eta_{IC}}\gamma = \frac{\frac{\partial}{\partial e}\phi(q_M|e^*)}{\phi(q_M|e^*)} \tag{54}$$

   Because of MLRP, the LHS of equation (54) is strictly increasing in q_m, and the RHS is strictly increasing in q_M. Thus, there is a continuum of $\{q_m, q_M\}$ that satisfy this equation such that for any pair in this continuum, q_M is strictly increasing in q_m.

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

$$-\gamma\rho \int_0^{q_m} q \frac{\partial}{\partial e}\phi(q|e^*)dq + \int_{q_m}^{\overline{q}} q \frac{\partial}{\partial e}\phi(q|e^*)dq = C'^*) \tag{55}$$

$$-\gamma\rho \int_0^{q_m} q\phi(q|e^*)dq + \int_{q_m}^{\overline{q}} q\phi(q|e^*)dq - C(e^*) = \overline{U} \tag{56}$$

The LHS of the IR in equation (56) is strictly decreasing in q_m. Thus, there exists at most one value of q_m such that equation (56) holds, and this value is strictly decreasing in $\overline{U}$. The derivative of the LHS of IC in equation (55) with respect to q_m is $q_m \frac{\partial}{\partial e}\phi(q_m|e^*)(-\gamma\rho - 1)$, 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 (56) 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:

$$-\gamma\rho \int_0^{q_m} q \frac{\partial}{\partial e}\phi(q|e^*)dq + \rho \int_{q_m}^{\overline{q}} q \frac{\partial}{\partial e}\phi(q|e^*)dq = C'^*) \tag{57}$$

$$-\gamma\rho \int_0^{q_m} q\phi(q|e^*)dq + \rho \int_{q_m}^{\overline{q}} q\phi(q|e^*)dq - C(e^*) = \overline{U} \tag{58}$$

The LHS of the IR in equation (58) is strictly decreasing in q_m. Thus, there exists at most one value of q_m such that equation (58) holds, and this value is strictly decreasing in $\overline{U}$. The derivative of the LHS of IC in equation (57) with respect to q_m is $q_m \frac{\partial}{\partial e}\phi(q_m|e^*)(-\gamma\rho - \rho)$, 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 (58) holds, and these values are independent of $\overline{U}$. In sum, generically we cannot have IC and IR binding with $q_M = \overline{q}$.

Proof of Corollary 1:

When the IC is binding with a contract as in Proposition 2, the IC can be written as in equation (52). In this equation, only the first term on the LHS depends on γ. Furthermore, with $q_m \leq q_0^{e^T}$ and a binding IC which implies $e^* = e^T$, we have $\frac{\partial}{\partial e}\phi(q|e^*) < 0$ for any $q < q_m$, so that the first term on the LHS of equation (52) is strictly positive. Moreover, with $q_m \leq q_0^{e^T}$ the IR must be binding as established in the proof of Proposition 2.

With a contract as in Proposition 2, the first derivatives of the LHS of the IC and IR in equations (52) and (53) with respect to γ are respectively:

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

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

With a contract as in Proposition 2, the first derivatives of the LHS of the IC in equation (52) with respect to q_m and q_M are respectively:

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

With a contract as in Proposition 2, the first derivatives of the LHS of the IR in equation (53) with respect to q_m and q_M are respectively:

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

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.

Proof of Proposition 3:

We describe the optimal contract when $e^T = 0$, i.e. the IC does not bind. 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) and (4) is simply $w(q) = 0$ for any q, so that, using equation (11) at any effort e^* with $\nu(0,q) \leq 0$ by assumption:

$$\int_0^{\overline{q}} u(0,q)\phi(q|e^*)dq = \int_0^{\overline{q}} \nu(0,q)\phi(q|e^*)dq \leq 0 \quad \Rightarrow \quad \int_0^{\overline{q}} u(0,q)\phi(q|e^*)dq - C(e^*) < \overline{U}$$

i.e. the PC is not satisfied, a contradiction.

The relaxed optimization problem with a nonbinding IC and a binding PC is:

$$\min_{w(q)} \int_0^{\overline{q}} w(q)\phi(q|e^*)dq \tag{59}$$

$$\text{s.t.} \quad \int_0^{\overline{q}} u(w(q),q)\phi(q|e^*)dq - C(e^*) = \overline{U} \tag{60}$$

$$0 \leq w(q) \leq q \tag{61}$$

Lemma 4 *Let the utility function be as in equation (17) and suppose that the IC is nonbinding. On any non-empty subinterval of* $[0,\overline{q}]$*, the optimal contract is such that* $w(q) \leq w^*(q)$*.*

Proof. This proof is by contradiction.

When equation (21) holds as an equality, it is easy to show that the optimal contract with a nonbinding IC is simply $w(q) \geq w^*(q)$ for all $q \in [0,\overline{q}]$.

Now consider the case when equation (21) is not satisfied as an equality. A contract such that $w(q) \geq w^*(q)$ for all $q \in [0,\overline{q}]$ would not solve the optimization program in equations (59)-(61). Indeed, due to equation (21) the agent would be strictly above his reservation utility so that the payment $w(q)$ could be reduced on some subinterval of $[0, \overline{q}]$, which would decrease the cost of the contract in equation (59) without violating the IR in equation (60) or the limited liability constraints in equation (61), a contradiction. Thus, any optimal contract is such that $w(q) < w^*(q)$ for some q when equation (21) is not satisfied as an equality. For a given contract, denote the subset of $[0, \overline{q}]$ such that $w(q) < w^*(q)$ by Q^-.

We now show that a contract with $w(q) > w^*(q)$ for any $q \in [0, \overline{q}]$ is suboptimal. Suppose that we have $w(q) \in (w^*(q), q]$ on a non-empty subinterval of $[0, \overline{q}]$, which we denote by Q^+. Consider the following perturbation for $q \in Q^-$, increase $w(q)$ by $\epsilon/\phi(q|e^*)$, and for q'^+ decrease $w(q')$ by $\epsilon/\phi(q'^*)$, where ϵ is positive and arbitrarily small. By construction, this perturbation is cost-neutral for a given effort, i.e. it does not change the principal's objective function. Now consider the effect on the LHS of the IR in equation (27). Since $w(q) \in [0, w^*(q))$ and $w(q'^*(q'), q']$, the change in the LHS of the IR is strictly positive since marginal utility is steeper at any point below the fair wage compared to any point above the fair wage (see equation (18)). Since the LHS of the IR increases and the IR is binding, standard arguments show that it is then possible to construct a contract that leaves the LHS of the 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 $w(q) > w^*(q)$ for some q.

■

By using Lemmas 5 and 6, which hold for $v'' \leq 0$ and $\nu'' \geq 0$, and Lemma 4 above, we get that the optimal contract when the IC is nonbinding takes the form:

$$w(q) = \begin{cases} 0 & \text{if } q < q_m \\ w^*(q) & \text{if } q \geq q_m \end{cases} \tag{62}$$

Proof of Proposition 4:

The first part of the proof shows that the agent optimally exerts a positive effort with the contract $w^*(q)$. By definition of e^F in equation (22):

$$\int_0^{\overline{q}} u(w^*(q), q) \frac{\partial}{\partial e} \phi(q|e^F) dq = C'^F) \tag{63}$$

We have $u(w^*(q), q) = v(w^*(q))$ by definition of the fair wage and the utility function, and $w^{*\prime}(q) > 0$ and $v' > 0$. Combining this with MLRP shows that the LHS of equation (63) is strictly positive. Finally, by assumption, $C'(0) = 0$ and $C'(e) > 0$ for any positive e, so that $e^F > 0$ when the LHS of equation (63) is strictly positive.

The second part of the proof shows by contradiction that the contract $w^*(q)$ is the least costly way of providing expected utility $\overline{U}$ to the agent while inducing effort e^F. If another contract $\hat{w}(q)$ induces the same effort and satisfies the IR at a smaller cost than $w^*(q)$, then we must have $\hat{w}(q) < w^*(q)$ for some $q \in \mathcal{Q}_-$ (otherwise this contract would not be associated with a lower cost), and $\hat{w}(q) > w^*(q)$ for some $q \in \mathcal{Q}_+$ given equation (21) satisfied as an equality (otherwise this contract would not provide expected utility $\overline{U}$). Now consider the agent's expected utility under contract $\hat{w}(q)$:

$$\int_0^{\overline{q}} u(\hat{w}(q), q)\phi(q|e^F)dq = \int_{[0,\overline{q}]\setminus(\mathcal{Q}_- \cup \mathcal{Q}_+)} u(w^*(q), q)\phi(q|e^F)dq$$
$$+ \int_{\mathcal{Q}_-} u(\hat{w}(q), q)\phi(q|e^F)dq + \int_{\mathcal{Q}_+} u(\hat{w}(q), q)\phi(q|e^F)dq \geq \overline{U}, \tag{64}$$

where the inequality is because the contract $\hat{w}(q)$ must satisfy the IR.

For $q \in \mathcal{Q}_-$, by the intermediate value theorem we know that for any given q there exists $\breve{w}(q) \in (\hat{w}(q), w^*(q))$ such that:

$$\begin{aligned} u(\hat{w}(q), q) &= u(w^*(q), q) + u'(\breve{w}(q), q)\left(\hat{w}(q) - w^*(q)\right) \\ &= u(w^*(q), q) + \nu'(\breve{w}(q), q)\left(\hat{w}(q) - w^*(q)\right) \end{aligned} \tag{65}$$

Integrating:

$$\int_{\mathcal{Q}_-} u(\hat{w}(q), q)\phi(q|e^F)dq = \int_{\mathcal{Q}_-} \left(u(w^*(q), q) + \nu'(\breve{w}(q), q)\left(\hat{w}(q) - w^*(q)\right)\right)\phi(q|e^F)dq$$

Let $\nu'_{\mathcal{Q}_-}$ be implicitly defined by:

$$\int_{\mathcal{Q}_-} u(\hat{w}(q), q)\phi(q|e^F)dq = \int_{\mathcal{Q}_-} u(w^*(q), q)\phi(q|e^F)dq + \nu'_{\mathcal{Q}_-} \int_{\mathcal{Q}_-} \left(\hat{w}(q) - w^*(q)\right)\phi(q|e^F)dq, \tag{66}$$

where $\nu'_{\mathcal{Q}_-} \geq \nu'^*(q), q)$ since $\nu'' \geq 0$.

Likewise, for $q \in \mathcal{Q}_+$, by the intermediate value theorem we know that for any given q there exists $\breve{w}(q) \in (w^*(q), \hat{w}(q))$ such that:

$$\begin{aligned} u(\hat{w}(q), q) &= u(w^*(q), q) - u'(\breve{w}(q), q)\left(\hat{w}(q) - w^*(q)\right) \\ &= u(w^*(q), q) - v'(\breve{w}(q), q)\left(\hat{w}(q) - w^*(q)\right) \end{aligned} \tag{67}$$

Integrating:

$$\int_{\mathcal{Q}_+} u(\hat{w}(q), q)\phi(q|e^F)dq = \int_{\mathcal{Q}_+} \left(u(w^*(q), q) + v'(\breve{w}(q), q)\left(\hat{w}(q) - w^*(q)\right)\right)\phi(q|e^F)dq$$

Let $v'_{\mathcal{Q}_+}$ be implicitly defined by:

$$\int_{\mathcal{Q}_+} u(\hat{w}(q), q)\phi(q|e^F)dq = \int_{\mathcal{Q}_+} u(w^*(q), q)\phi(q|e^F)dq + v'_{\mathcal{Q}_+} \int_{\mathcal{Q}_+} \left(\hat{w}(q) - w^*(q)\right)\phi(q|e^F)dq, \tag{68}$$

where $v'_{\mathcal{Q}_+} \leq v'^*(q), q)$ since $v'' \leq 0$.

In sum, substituting in the RHS of equation (64):

$$
\begin{aligned}
&\int_{[0,\overline{q}]\setminus(\mathcal{Q}_- \cup \mathcal{Q}_+)} u(w^*(q), q)\phi(q|e^F)dq + \int_{\mathcal{Q}_-} u(\hat{w}(q), q)\phi(q|e^F)dq + \int_{\mathcal{Q}_+} u(\hat{w}(q), q)\phi(q|e^F)dq \\
&= \int_{[0,\overline{q}]\setminus(\mathcal{Q}_- \cup \mathcal{Q}_+)} u(w^*(q), q)\phi(q|e^F)dq \\
&+ \int_{\mathcal{Q}_-} u(w^*(q), q)\phi(q|e^F)dq + \nu'_{\mathcal{Q}_-} \int_{\mathcal{Q}_-} (\hat{w}(q) - w^*(q))\, \phi(q|e^F)dq \\
&+ \int_{\mathcal{Q}_+} u(w^*(q), q)\phi(q|e^F)dq + v'_{\mathcal{Q}_+} \int_{\mathcal{Q}_+} (\hat{w}(q) - w^*(q))\, \phi(q|e^F)dq \\
&= \int_0^{\overline{q}} u(w^*(q), q)\phi(q|e^F)dq + \nu'_{\mathcal{Q}_-} \int_{\mathcal{Q}_-} (\hat{w}(q) - w^*(q))\, \phi(q|e^F)dq + v'_{\mathcal{Q}_+} \int_{\mathcal{Q}_+} (\hat{w}(q) - w^*(q))\, \phi(q|e^F)dq,
\end{aligned} \tag{69}
$$

where $v'_{\mathcal{Q}_+} < \nu'_{\mathcal{Q}_+}$ due to the assumptions on the first derivatives with respect to w of the function v and ν. Since the contract $\hat{w}(q)$ must provide expected utility of at least $\overline{U}$ to satisfy IR, equations (21) satisfied as an equality and (69) imply:

$$
\nu'_{\mathcal{Q}_-} \int_{\mathcal{Q}_-} \underbrace{(\hat{w}(q) - w^*(q))}_{<0} \phi(q|e^F)dq + v'_{\mathcal{Q}_+} \int_{\mathcal{Q}_+} \underbrace{(\hat{w}(q) - w^*(q))}_{>0} \phi(q|e^F)dq \geq 0 \tag{70}
$$

The inequalities under braces in equation (70) combined with $v'_{\mathcal{Q}_+} < \nu'_{\mathcal{Q}_+}$ give:

$$
\begin{aligned}
&\int_{\mathcal{Q}_+} \underbrace{(\hat{w}(q) - w^*(q))}_{>0} \phi(q|e^F)dq \geq \int_{\mathcal{Q}_-} \underbrace{(w^*(q) - \hat{w}(q))}_{>0} \phi(q|e^F)dq \\
&\Leftrightarrow \quad \int_{\mathcal{Q}_- \cup \mathcal{Q}_+} \hat{w}(q)\phi(q|e^F)dq \geq \int_{\mathcal{Q}_- \cup \mathcal{Q}_+} w^*(q)\phi(q|e^F)dq
\end{aligned} \tag{71}
$$

By definition of the subsets $\mathcal{Q}_-$ and $\mathcal{Q}_+$, this implies that the expected cost of the contract $\hat{w}(q)$ is higher than the expected cost of the contract $w^*(q)$. This concludes the second part of the proof.

Proof of Proposition 5:

By definition of the fair wage, the (local) sensitivity of the fair wage to output is:

$$
\frac{dw^*(q)}{dq} = -\frac{\frac{\partial}{\partial q}\{v(w^*(q)) - \nu(w^*(q), q)\}}{\frac{\partial}{\partial w}\{v(w^*(q)) - \nu(w^*(q), q)\}} = \frac{\nu'_q(w^*(q), q)}{v'(w^*(q)) - \nu'_w(w^*(q), q)}.
$$

To ensure that the fair wage is increasing in output but remains lower than output (i.e. the principal limited liability constraint would not be violated if the agent were paid a fair wage), we restrict attention to parameter values such that:

$$w^*(0) = 0 \quad \text{and} \quad \frac{\nu'_q(w^*(q), q)}{v'(w^*(q)) - \nu'_w(w^*(q), q)} \in (0, 1) \;\; \forall q. \tag{72}$$

The local convexity of the fair wage with respect to output is:

$$\begin{aligned} \frac{d^2 w^*(q)}{dq^2} = & \frac{\left(w^{*\prime}(q)\nu''^*_{wq}(q), q) + \nu''^*_{qq}(q), q)\right)(v'^*(q)) - \nu'^*_w(q), q))}{(v'^*(q)) - \nu'^*_w(q), q))^2} \\ & - \frac{\nu'^*_q(q), q)(w^{*\prime}(q)v''^*(q)) - w^{*\prime}(q)\nu''^*_{ww}(q), q) - \nu''^*_{wq}(q), q))}{(v'^*(q)) - \nu'^*_w(q), q))^2} \leq 0, \end{aligned} \tag{73}$$

where the inequality is due to $w^{*\prime}(q) > 0$, $\nu'_q < 0$, $\nu''_{wq} = 0$, $\nu''_{qq} = 0$, $v'' \leq 0$, and $\nu''_{ww} \geq 0$, and the inequality is strict with either $v'' < 0$ or $\nu''_{ww} > 0$. With equation (73), we have:

$$w^*(q) \leq w^*(0) + w^{*\prime}(0)q, \tag{74}$$

where $w^*(0) = 0$, see equation (72). Given the contract $w^*(q)$, the principal's objective function is strictly increasing in effort for any e, all else equal:

$$\begin{aligned} \frac{\partial}{\partial e}\int_0^{\overline{q}} (q - w^*(q))\, \phi(q|e)dq &= \int_0^{\overline{q}} (q - w^*(q))\, \frac{\partial}{\partial e}\phi(q|e)dq \\ &\geq \left(1 - w^{*\prime}(0)\right)\int_0^{\overline{q}} q\frac{\partial}{\partial e}\phi(q|e)dq \\ &\geq 0, \end{aligned} \tag{75}$$

where the first inequality used equation (74), and the second inequality used the assumption in equation (72).

When reservation utility $\overline{U}$ is as in equation (21) satisfied as an equality, a contract $w^*(q)$ is optimal when $e^T = 0$. Moreover, with a contract $w^*(q)$, by definition of e^F in equation (22), the agent optimally exerts effort e^F. For incentive compatibility, the contract must deviate from $w^*(q)$ for at least some outputs while maintaining the agent's expected utility at the same level when effort is e_1 instead of e^F. Consider the agent's expected utility under a given contract $\hat{w}(q)$ that elicits effort e_1 and satisfies the IR:

$$\begin{aligned} \overline{U} = \int_0^{\overline{q}} u(w^*(q), q)\phi(q|e^F)dq \leq & \int_{[0,\overline{q}]\setminus(\mathcal{Q}_- \cup \mathcal{Q}_+)} u(w^*(q), q)\phi(q|e_1)dq \\ + \int_{\mathcal{Q}_-} u(\hat{w}(q), q)\phi(q|e_1)dq + & \int_{\mathcal{Q}_+} u(\hat{w}(q), q)\phi(q|e_1)dq. \end{aligned} \tag{76}$$

Substituting from equations (66) and (68) in the RHS of equation (76):

$$\begin{aligned}
&\int_{[0,\overline{q}]\setminus(\mathcal{Q}_-\cup\mathcal{Q}_+)} u(w^*(q),q)\phi(q|e_1)dq + \int_{\mathcal{Q}_-} u(\hat{w}(q),q)\phi(q|e_1)dq + \int_{\mathcal{Q}_+} u(\hat{w}(q),q)\phi(q|e_1)dq \\
&= \int_{[0,\overline{q}]\setminus(\mathcal{Q}_-\cup\mathcal{Q}_+)} u(w^*(q),q)\phi(q|e_1)dq \\
&+ \int_{\mathcal{Q}_-} u(w^*(q),q)\phi(q|e_1)dq + \nu'_{\mathcal{Q}_-}\int_{\mathcal{Q}_-} \left(\hat{w}(q)-w^*(q)\right)\phi(q|e_1)dq \\
&+ \int_{\mathcal{Q}_+} u(w^*(q),q)\phi(q|e_1)dq + v'_{\mathcal{Q}_+}\int_{\mathcal{Q}_+} \left(\hat{w}(q)-w^*(q)\right)\phi(q|e_1)dq \\
&= \int_0^{\overline{q}} u(w^*(q),q)\phi(q|e_1)dq + \nu'_{\mathcal{Q}_-}\int_{\mathcal{Q}_-} \left(\hat{w}(q)-w^*(q)\right)\phi(q|e_1)dq + v'_{\mathcal{Q}_+}\int_{\mathcal{Q}_+} \left(\hat{w}(q)-w^*(q)\right)\phi(q|e_1)dq,
\end{aligned} \tag{77}$$

where $v'_{\mathcal{Q}_+} < \nu'_{\mathcal{Q}_+}$ due to the assumptions on the first derivatives with respect to w of the function v and ν. Since the contract $\hat{w}(q)$ must provide expected utility of at least $\overline{U}$ to satisfy IR, equations (21) satisfied as an equality and (77) imply:

$$\nu'_{\mathcal{Q}_-}\int_{\mathcal{Q}_-} \underbrace{\left(\hat{w}(q)-w^*(q)\right)}_{<0}\phi(q|e_1)dq + v'_{\mathcal{Q}_+}\int_{\mathcal{Q}_+} \underbrace{\left(\hat{w}(q)-w^*(q)\right)}_{>0}\phi(q|e_1)dq \geq 0 \tag{78}$$

The inequalities under braces in equation (78) combined with $v'_{\mathcal{Q}_+} < \nu'_{\mathcal{Q}_+}$ give:

$$\begin{aligned}
&\int_{\mathcal{Q}_+} \underbrace{\left(\hat{w}(q)-w^*(q)\right)}_{>0}\phi(q|e_1)dq \geq \int_{\mathcal{Q}_-} \underbrace{\left(w^*(q)-\hat{w}(q)\right)}_{>0}\phi(q|e_1)dq \\
&\qquad\Leftrightarrow \quad \int_{\mathcal{Q}_-\cup\mathcal{Q}_+} \hat{w}(q)\phi(q|e_1)dq \geq \int_{\mathcal{Q}_-\cup\mathcal{Q}_+} w^*(q)\phi(q|e_1)dq
\end{aligned} \tag{79}$$

By definition of the subsets $\mathcal{Q}_-$ and $\mathcal{Q}_+$, this implies:

$$\mathbb{E}[\hat{w}(q)|e_1] > \mathbb{E}[w^*(q)|e_1]. \tag{80}$$

This in turn implies:

$$\mathbb{E}[q-\hat{w}(q)|e_1] < \mathbb{E}[q-w^*(q)|e_1]. \tag{81}$$

Finally, since the principal's objective function is increasing in effort, ceteris paribus, given contract $w^*(q)$ (see equation (75)), for $e_1 < e^F$ we have:

$$\mathbb{E}[q-w^*(q)|e^F] \geq \mathbb{E}[q-w^*(q)|e_1]. \tag{82}$$

Combining equations (81) and (82):

$$\mathbb{E}[q - w^*(q)|e^F] > \mathbb{E}[q - \hat{w}(q)|e_1].$$

This shows that the principal's objective function is higher when inducing effort e^F than when inducing effort e_1.

Proof of Proposition 6:

By Lemma 1, when the condition in equation (7) holds, we can replace the IC in equation (2) by the FOC:

$$\int_0^{\overline{q}} u(w(q), q)\frac{\partial}{\partial e}\phi(q|e^*)dq = C'^*). \tag{83}$$

The proof has two steps.

First step

Lemma 5 *For any q and q' such that $q > q'$ (without loss of generality, "WLOG"), the optimal contract is such that $w(q) \geq w(q')$.*

Proof. There are three possible cases to consider. In each case, the proof is by contradiction.

First case. Suppose that the optimal contract is such that there exists $q > q'$ with $q = q' + \varepsilon$ and $\varepsilon \to 0$ such that $w(q) < w(q')$ where $w(q) \in [0, w^*(q)]$ and $w(q'^*(q')]$. If $w(q) = w^*(q)$ or if $w(q') = 0$, we automatically have $w(q) \geq w(q')$. Now consider cases where $w(q) < w^*(q)$ and $w(q') > 0$.

There are two cases. First, define $\Delta \equiv w^*(q) - w(q)$, and suppose that there exists ϵ such that $w(q') - \Delta\epsilon \geq 0$ and:

$$[\nu(w^*(q)) - \nu(w(q))]\,\phi(q|e^*) + [\nu(w(q') - \Delta\epsilon) - \nu(w(q'))]\,\phi(q'^*) = 0 \tag{84}$$

$$\Leftrightarrow \quad \frac{\nu(w^*(q)) - \nu(w(q))}{\nu(w(q')) - \nu(w(q') - \Delta\epsilon)} = \frac{\phi(q'^*)}{\phi(q|e^*)} \tag{85}$$

Then set $w_1(q) = w^*(q)$, and $w_1(q') = w(q') - \Delta\epsilon$ (i.e. the variable ϵ is chosen so that the switch to this new contract w_1 does not change the LHS of the IR). Second, suppose that there does not exist ϵ such that $w_1(q') \geq 0$ and equation (85) holds, i.e.

$$[\nu(w^*(q)) - \nu(w(q))]\,\phi(q|e^*) + [\nu(0) - \nu(w(q'))]\,\phi(q'^*) > 0. \tag{86}$$

Then set $w_1(q') = 0$, define $\Delta \equiv w_1(q')$ and $w_1(q) = w(q) + \Delta/\epsilon$, where ϵ is set so that (comparing equations (86) and (87) shows that ϵ exists since we must have $w(q) + \Delta/\epsilon < w^*(q)$, and $w(q) <$

$w^*(q)$ in the case considered here):

$$[\nu(w(q)+\Delta/\epsilon)-\nu(w(q))]\,\phi(q|e^*)+[\nu(0)-\nu(w(q'))]\,\phi(q'^*)=0 \tag{87}$$

$$\Leftrightarrow \quad \frac{\nu(w(q)+\Delta/\epsilon)-\nu(w(q))}{\nu(w(q'))-\nu(0)}=\frac{\phi(q'^*)}{\phi(q|e^*)} \quad \Leftrightarrow \tag{88}$$

In any case, we have:

$$w_1(q') \leq w(q) < w(q') < w_1(q) \tag{89}$$

Indeed, we have $w(q) < w(q')$ by assumption. Moreover, in the first case above, we have $w_1(q) = w^*(q)$, which is strictly larger than $w^*(q')$ (since $q' < q$ and $w^*(q)$ is strictly increasing in q), and therefore strictly larger than $w(q')$. Finally, given that we have $w(q) < w(q') < w_1(q)$ and $\frac{\phi(q'^*)}{\phi(q|e^*)} \to 1$ with a continuously differentiable density and $q' \to q$, the first inequality in equation (89) is necessary for equation (84) to hold. In the second case above, we have $w_1(q') = 0$, i.e. $w_1(q') \leq w(q)$. Finally, given that we have $w_1(q') \leq w(q) < w(q')$ and $\frac{\phi(q'^*)}{\phi(q|e^*)} \to 1$ with a continuously differentiable density and $q' \to q$, the last inequality in equation (89) is necessary for equation (87) to hold.

There are again two cases. First, if $\frac{\partial}{\partial e}\phi(q|e^*) > 0$, then equation (85) or (88) and MLRP imply:

$$\frac{\nu(w_1(q))-\nu(w(q))}{\nu(w(q'))-\nu(w_1(q'))}=\frac{\phi(q'^*)}{\phi(q|e^*)}>\frac{\frac{\partial}{\partial e}\phi(q'^*)}{\frac{\partial}{\partial e}\phi(q|e^*)} \tag{90}$$

$$\Rightarrow \qquad [\nu(w_1(q))-\nu(w(q))]\,\frac{\partial}{\partial e}\phi(q|e^*)+[\nu(w_1(q'))-\nu(w(q'))]\,\frac{\partial}{\partial e}\phi(q'^*)>0 \tag{91}$$

Second, if $\frac{\partial}{\partial e}\phi(q|e^*) < 0$, then equation (85) or (88) and MLRP imply:

$$\frac{\nu(w_1(q))-\nu(w(q))}{\nu(w(q'))-\nu(w_1(q'))}=\frac{\phi(q'^*)}{\phi(q|e^*)}<\frac{\frac{\partial}{\partial e}\phi(q'^*)}{\frac{\partial}{\partial e}\phi(q|e^*)} \tag{92}$$

$$\Rightarrow \qquad [\nu(w_1(q))-\nu(w(q))]\,\frac{\partial}{\partial e}\phi(q|e^*)+[\nu(w_1(q'))-\nu(w(q'))]\,\frac{\partial}{\partial e}\phi(q'^*)>0 \tag{93}$$

That is, in both cases the IC is relaxed.

Switching to the new contract w_1 decreases the expected cost to the principal if and only if:

$$[w_1(q)-w(q)]\,\phi(q|e^*)+[w_1(q')-w(q')]\,\phi(q'^*)\leq 0 \quad \Leftrightarrow \quad \frac{\phi(q'^*)}{\phi(q|e^*)}\geq\frac{1}{\epsilon} \tag{94}$$

By MLRP, $\frac{\frac{\partial}{\partial e}\phi(q'^*)}{\phi(q'^*)} < \frac{\frac{\partial}{\partial e}\phi(q|e^*)}{\phi(q|e^*)}$ since $q > q'$ by assumption. In addition, since $\nu'' \geq 0$ and equation (89) holds:

$$\frac{\nu(w_1(q))-\nu(w(q))}{\nu(w(q'))-\nu(w_1(q'))}\geq\frac{1}{\epsilon}, \tag{95}$$

with a strict inequality for $\nu'' > 0$. Combining equations in either (90) or (92) and (95), we get:

$$\frac{\phi(q'^*)}{\phi(q|e^*)} = \frac{\nu(w_1(q)) - \nu(w(q))}{\nu(w(q')) - \nu(w_1(q'))} \geq \frac{1}{\epsilon},$$

with a strict inequality for $\nu'' > 0$. This establishes that equation (94) holds, i.e. the principal's objective function for a given effort level is improved (strictly if $\nu'' > 0$) while the LHS of the relaxed IC in equation (83) increases at a given effort level. With the FOA, the agent's problem is concave in effort, so that the agent will optimally exert higher effort than with the initial contract, which increases the principal's objective function. The IR remains satisfied (if it is satisfied at effort e_1 but the agent is better off under effort e_2, then it is satisfied at effort e_2 as well.). This contradicts the optimality of the initial contract.

Second case. Suppose that the optimal contract is such that there exists $q > q'$ with $q = q' + \varepsilon$ and $\varepsilon \to 0$ such that $w(q) < w(q')$ where $w(q) \in [w^*(q), q]$ and $w(q'^*(q'), q']$. If $w(q) = q$ or if $w(q'^*(q')$, we automatically have $w(q) \geq w(q')$. Now consider cases where $w(q) < q$ and $w(q'^*(q')$. Set $w_1(q) = w(q) + \Delta$, where Δ is positive and arbitrarily small, and $w_1(q') = w(q') - \Delta\epsilon$, with $\epsilon > 0$. The variable ϵ is chosen so that the switch to this new contract w_1 does not change the LHS of the IR:

$$\begin{aligned}&[v(w_1(q)) - v(w(q))]\,\phi(q|e^*) + [v(w_1(q')) - v(w(q'))]\,\phi(q'^*) = 0\\ &\qquad\Leftrightarrow \quad \frac{v(w_1(q)) - v(w(q))}{v(w_1(q')) - v(w(q'))} = \frac{\phi(q'^*)}{\phi(q|e^*)}\end{aligned} \tag{96}$$

The variable Δ is chosen so that we have: $w(q) < w_1(q) < w_1(q') < w(q')$. When $\frac{\partial}{\partial e}\phi(q|e^*) > 0$, equation (96) and MLRP give:

$$\begin{aligned}&\frac{\frac{\partial}{\partial e}\phi(q'^*)}{\frac{\partial}{\partial e}\phi(q|e^*)} < \frac{\phi(q'^*)}{\phi(q|e^*)} = \frac{v(w_1(q)) - v(w(q))}{v(w(q')) - v(w_1(q'))}\\ &\qquad\Leftrightarrow \quad [v(w_1(q)) - v(w(q))]\,\frac{\partial}{\partial e}\phi(q|e^*) + [v(w_1(q')) - v(w(q'))]\,\frac{\partial}{\partial e}\phi(q'^*) > 0.\end{aligned} \tag{97}$$

When $\frac{\partial}{\partial e}\phi(q|e^*) < 0$, equation (96) and MLRP give:

$$\begin{aligned}&\frac{\frac{\partial}{\partial e}\phi(q'^*)}{\frac{\partial}{\partial e}\phi(q|e^*)} > \frac{\phi(q'^*)}{\phi(q|e^*)} = \frac{v(w_1(q)) - v(w(q))}{v(w(q')) - v(w_1(q'))}\\ &\qquad\Rightarrow \quad [v(w_1(q)) - v(w(q))]\,\frac{\partial}{\partial e}\phi(q|e^*) + [v(w_1(q')) - v(w(q'))]\,\frac{\partial}{\partial e}\phi(q'^*) > 0.\end{aligned} \tag{98}$$

That is, in both cases the IC is relaxed.

Switching to the new contract w_1 decreases the expected cost to the principal if and only if:

$$[w_1(q) - w(q)]\,\phi(q|e^*) + [w_1(q') - w(q')]\,\phi(q'^*) \leq 0 \quad \Leftrightarrow \quad \frac{\phi(q'^*)}{\phi(q|e^*)} \geq \frac{1}{\epsilon} \tag{99}$$

By MLRP, $\frac{\frac{\partial}{\partial e}\phi(q'^*)}{\phi(q'^*)} < \frac{\frac{\partial}{\partial e}\phi(q|e^*)}{\phi(q|e^*)}$ since $q > q'$ by assumption. In addition, since we have $w(q) < w_1(q) < w_1(q') < w(q')$ and $v'' \leq 0$:

$$\frac{v(w_1(q)) - v(w(q))}{v(w(q')) - v(w_1(q'))} \geq \frac{1}{\epsilon},$$

with a strict inequality for $v'' < 0$. Combining the inequalities above, we get:

$$\frac{\phi(q'^*)}{\phi(q|e^*)} = \frac{v(w_1(q)) - v(w(q))}{v(w(q')) - v(w_1(q'))} \geq \frac{1}{\epsilon},$$

with a strict inequality for $v'' < 0$. This establishes that equation (99) holds, i.e. the principal's objective function for a given effort level is improved (strictly if $v'' < 0$) while the LHS of the relaxed IC in equation (83) increases at a given effort level. With the FOA, the agent's problem is concave in effort, so that the agent will optimally exert higher effort than with the initial contract, which increases the principal's objective function. The IR remains satisfied (if it is satisfied at effort e_1 but the agent is better off under effort e_2, then it is satisfied at effort e_2 as well.). This contradicts the optimality of the initial contract.

Third case. Suppose that the optimal contract is such that there exists $q > q'$ with $q = q' + \varepsilon$ and $\varepsilon \to 0$ such that $w(q) < w(q')$ where $w(q) \in [0, w^*(q)]$ and $w(q'^*(q'), q']$. If $w(q) = w^*(q)$, we automatically have $w(q) \geq w(q')$. Now consider cases where $w(q) < w^*(q)$. Set $w_1(q) = w(q) + \Delta$, where Δ is positive and arbitrarily small, and $w_1(q') = w(q') - \Delta\epsilon$, with $\epsilon > 0$. The variable ϵ is chosen so that the switch to this new contract w_1 does not change the LHS of the IR:

$$[\nu(w_1(q)) - \nu(w(q))]\,\phi(q|e^*) + [v(w_1(q')) - v(w(q'))]\,\phi(q'^*) = 0$$
$$\Leftrightarrow \quad \frac{\nu(w_1(q)) - \nu(w(q))}{v(w_1(q')) - v(w(q'))} = \frac{\phi(q'^*)}{\phi(q|e^*)} \tag{100}$$

The variable Δ must be sufficiently small for $w_1(q) \leq w^*(q)$ and $w_1(q'^*(q')$. When $\frac{\partial}{\partial e}\phi(q|e^*) > 0$, equation (100) and MLRP give:

$$\frac{\frac{\partial}{\partial e}\phi(q'^*)}{\frac{\partial}{\partial e}\phi(q|e^*)} < \frac{\phi(q'^*)}{\phi(q|e^*)} = \frac{\nu(w_1(q)) - \nu(w(q))}{v(w(q')) - v(w_1(q'))}$$
$$\Rightarrow \qquad [\nu(w_1(q)) - \nu(w(q))]\,\frac{\partial}{\partial e}\phi(q|e^*) + [v(w_1(q')) - v(w(q'))]\,\frac{\partial}{\partial e}\phi(q'^*) > 0. \tag{101}$$

When $\frac{\partial}{\partial e}\phi(q|e^*) < 0$, equation (100) and MLRP give:

$$\frac{\frac{\partial}{\partial e}\phi(q'^*)}{\frac{\partial}{\partial e}\phi(q|e^*)} > \frac{\phi(q'^*)}{\phi(q|e^*)} = \frac{\nu(w_1(q)) - \nu(w(q))}{v(w(q')) - v(w_1(q'))}$$
$$\Rightarrow \qquad [\nu(w_1(q)) - \nu(w(q))]\,\frac{\partial}{\partial e}\phi(q|e^*) + [v(w_1(q')) - v(w(q'))]\,\frac{\partial}{\partial e}\phi(q'^*) > 0. \qquad (102)$$

That is, in both cases the IC is relaxed.

Switching to the new contract w_1 decreases the expected cost to the principal if and only if:

$$[w_1(q) - w(q)]\,\phi(q|e^*) + [w_1(q') - w(q')]\,\phi(q'^*) < 0 \quad \Leftrightarrow \quad \frac{\phi(q'^*)}{\phi(q|e^*)} > \frac{1}{\epsilon} \qquad (103)$$

By MLRP, $\frac{\frac{\partial}{\partial e}\phi(q'^*)}{\phi(q'^*)} < \frac{\frac{\partial}{\partial e}\phi(q|e^*)}{\phi(q|e^*)}$ since $q > q'$ by assumption. In addition, since $\nu'' \geq 0$, $v'' \leq 0$, $\lim_{w\searrow 0}\nu'(w,q) > \lim_{w\searrow 0} v'(w)$:

$$\frac{\nu(w_1(q)) - \nu(w(q))}{v(w(q')) - v(w_1(q'))} \geq \frac{1}{\epsilon},$$

with a strict inequality for $\nu'' > 0$ and $v'' < 0$. Combining the inequalities above, we get:

$$\frac{\phi(q'^*)}{\phi(q|e^*)} = \frac{\nu(w_1(q)) - \nu(w(q))}{v(w(q')) - v(w_1(q'))} \geq \frac{1}{\epsilon},$$

with a strict inequality for $\nu'' > 0$ and $v'' < 0$. This establishes that equation (103) holds, i.e. the principal's objective function for a given effort level is improved (strictly if $\nu'' > 0$ and $v'' < 0$) while the LHS of the relaxed IC in equation (83) increases at a given effort level. With the FOA, the agent's problem is concave in effort, so that the agent will optimally exert higher effort than with the initial contract, which increases the principal's objective function. The IR remains satisfied (if it is satisfied at effort e_1 but the agent is better off under effort e_2, then it is satisfied at effort e_2 as well.). This contradicts the optimality of the initial contract. ■

Lemma 6 *On any non-empty subinterval of* $[0, \overline{q}]$*, we have* $w(q) \notin (0, w^*(q))$*.*

Proof. An optimal contract is nondecreasing in q according to Lemma 5. Accordingly, consider any given initial incentive-compatible contract with $q > q'$, $w(q) \geq w(q')$, $w(q) \in (0, w^*(q))$, $w(q'^*(q'))$, and the following perturbation: increase $w(q)$ by $\epsilon/\phi(q|e^*)$, and decrease $w(q')$ by $\epsilon/\phi(q'^*)$, where ϵ is positive and arbitrarily small. By construction, this perturbation is cost-neutral for a given effort, i.e. it does not change the principal's objective function. Now consider the effect on the LHS of the IR in (3). With $\nu'' \geq 0$ and $w(q) \geq w(q')$, we have $\nu'(w(q)) \geq \nu'(w(q'))$, so that the change in the LHS of the IR in (3) is:

$$\frac{\epsilon}{\phi(q|e^*)}\nu'^*) - \frac{\epsilon}{\phi(q'^*)}\nu'(w(q'))\phi(q'^*) \geq 0.$$

Now consider the effect on the LHS of the IC in (83) with a utility function as in equation (17). The change in the LHS of the IC for this perturbation is:

$$\epsilon\left(\nu'^{*}) - \nu'(w(q'))LR(q'^{*})\right)$$

By MLRP we have $LR(q|e^*) > LR(q'^*)$. In addition, $w(q) \geq w(q')$ so that $\nu'(w(q)) \geq \nu'(w(q'))$ due to $\nu'' \geq 0$. Thus, the LHS of the IC increases. In addition, given that the agent's problem is concave in effort given assumptions on the FOA, the agent will optimally exert higher effort than with the initial contract, which increases the principal's objective function. The IR remains satisfied (if it is satisfied at effort e_1 but the agent is better off under effort e_2, then it is satisfied at effort e_2 as well). This contradicts the optimality of the initial contract.

■

Second step

For now we ignore the limited liability constraints in equation (4). The FOC with respect to $w(q)$ in the principal's optimization program in equations (1)-(3) is:

$$\phi(q|e^*) - \lambda_{IR}u'^{*}_{w}) - \lambda_{IC}u'_w(w(q),q)\frac{\partial}{\partial e}\phi(q|e^*) = 0$$

$$\Leftrightarrow \quad \frac{1}{u'_w(w(q),q)} = \lambda_{IR} + \lambda_{IC}\frac{\frac{\partial}{\partial e}\phi(q|e^*)}{\phi(q|e^*)}. \tag{104}$$

Combining Lemma 5 and Lemma 6, we have $w(q) = 0$ for $q \in [0, q_m]$, for some $q_m \in [0, \overline{q}]$. For now, consider a given $q_m \in [0, \overline{q}]$. Due to results in the first step (including that $w(q)$ is nondecreasing in q and $w(q) \notin (0, w^*(q))$ for any q) and principal limited liability, for $q > q_m$ the payment is $w(q) \in [w^*(q), q]$ so that $u(w,q) = v(w)$ for $q > q_m$. Indeed, using the specification of the utility function in equation (17), we have $u_w(w^*(q),q) = v'^{*}(q))$ by definition of the fair wage, and $u_w(w(q),q) = v'(w(q))$ for any $w(q) > w^*(q)$ due to the assumption in equation (72). That is, for a given q_m, the optimization problem that gives the optimal contract to induce effort e^* can be rewritten as:

$$\min_{w(q)} \int_{q_m}^{\overline{q}} w(q)\phi(q|e^*)dq \tag{105}$$

$$\text{s.t.} \quad \int_0^{q_m} u(0,q)\frac{\partial}{\partial e}\phi(q|e^*)dq + \int_{q_m}^{\overline{q}} v(w(q))\frac{\partial}{\partial e}\phi(q|e^*)dq = C'^{*}) \tag{106}$$

$$\int_0^{q_m} u(0,q)\phi(q|e^*)dq + \int_{q_m}^{\overline{q}} v(w(q))\phi(q|e^*)dq \geq \overline{U} \tag{107}$$

$$w(q) \in [w^*(q), q] \; \forall q \tag{108}$$

Using the notation in Jewitt, Kadan, and Swinkels (2008), we have $\underline{m}(q) = w^*(q)$ and $\overline{m}(q) = q$, and the reasoning in Proposition 1 of their paper applies (note that the first terms on the LHS of equations (106) and (107) are independent of $w(q)$), so that the optimal contract is defined implicitly by:

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

with $\lambda_{IR} \ge 0$ and $\lambda_{IC} > 0$, and these Lagrange multipliers depend on q_m.